

University of San Diego

Digital USD

Dissertations

Theses and Dissertations

2013-05-01

Performance Assessment for Quality Teaching: Three Critical Variables for Measuring and Improving Teaching and Learning

Kathleen L. Gallagher PhD
University of San Diego

Follow this and additional works at: <https://digital.sandiego.edu/dissertations>



Part of the [Leadership Studies Commons](#)

Digital USD Citation

Gallagher, Kathleen L. PhD, "Performance Assessment for Quality Teaching: Three Critical Variables for Measuring and Improving Teaching and Learning" (2013). *Dissertations*. 842.
<https://digital.sandiego.edu/dissertations/842>

This Dissertation: Open Access is brought to you for free and open access by the Theses and Dissertations at Digital USD. It has been accepted for inclusion in Dissertations by an authorized administrator of Digital USD. For more information, please contact digital@sandiego.edu.

**PERFORMANCE ASSESSMENT FOR QUALITY TEACHING:
THREE CRITICAL VARIABLES FOR MEASURING AND IMPROVING
TEACHING AND LEARNING**

By

Kathleen L. Gallagher

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy
University of San Diego

May 2013

Dissertation Committee

Fred J. Galloway, Ed.D., Chair
Zachary Green, Ph.D., Member
Rose Linda P. Martinez, Ed.D., Member
Sid Salazar, Ed.D., Member
Ronald Heck, Ph.D., Member

© Copyright by Kathleen L. Gallagher
All Rights Reserved 2013

ABSTRACT

While outstanding teachers are any school system's most important investment, assessing the quality of instructional practice has proven to be an ongoing challenge for the profession. Despite assertions that effective teachers are the single most important school-related factor responsible for increased learning, no teacher's employment is dependent on their performance in the classroom or the quality of instruction that they provide. This problem has fueled a growing mistrust in school districts nationwide, specifically in the area of teacher evaluation. One possible explanation is that the profession lacks the scaled level of expertise needed to evaluate instruction consistently and in a manner that effectively informs the improvement process.

In an effort to both strengthen the teacher evaluation process and significantly improve the quality of instruction in classrooms, an observation instrument was developed which measures the critical skills associated with highly effective teaching. These include a teacher's content knowledge, pedagogical expertise, and the ability to establish a classroom culture conducive to sustained learning. The instrument is also consistent with the new Common Core State Standards, and defines quality as the level at which a teacher facilitates multi-directional interactions with the class that result in authentic cognitive engagement and increased subject matter competence. An important distinguishing factor is the instrument's focus on student outcomes related to participation, critical thinking, and academic language as opposed to traditional observations, which focus on teacher behaviors.

The study had both quantitative and qualitative components. Multilevel modeling techniques were used to examine the effects of instructional quality on student growth

trajectories in English and math in two California middle schools. The effects proved both positive and significant in both subject areas, but particularly in mathematics where one standard deviation of instructional quality produced an 11-point gain on the California Standards Test. A cross-case narrative analysis also identified the actions taken by teachers that resulted in the highest and lowest levels of instructional quality. Contributions of this study include an efficient model for evaluating instructional effectiveness, methods for informing and differentiating professional development, and an increased understanding of whether or not all students have access to high quality instruction.

DEDICATION

To Edward, Katherine, and Paul

and all of the children in the world who deserve high quality teachers.

To Dr. Martin Luther King, Jr.

Whose words to his own children remain a steadfast source of inspiration in my life:

“I’m going to work and do everything that I can do to see that you get a good education. I don’t ever want you to forget that there are millions of God’s children who will not and cannot get a good education, and I don’t want you feeling that you are better than they are. For you will never be what you ought to be until they are what they ought to be.”

ACKNOWLEDGEMENTS

There are many people who have contributed to the work that is presented here. First and foremost, I salute all of the educators who take their work seriously enough to sit down and write about it. Without them, I would never have learned what it means to be a high quality teacher. I appreciate Lucy McCormick Calkins who taught me the art of teaching reading and writing, Marilyn Burns, who transformed my understanding of what it means to teach mathematics, and Dr. Ronald Heck who taught me how to conduct research in education. I also am deeply indebted to all of the teachers who have helped frame this analysis. Without their input and feedback, this study would not have been possible.

Each of my dissertation committee members has inspired my work in unique and highly significant ways. My chair, Dr. Fred Galloway, believed from the start that my efforts had value and encouraged me without hesitation. Dr. Zachary Green pushed my own development with compassion and intensity. Dr. Rose Martinez served as a constant reminder that great teaching is an art form that requires creativity and emotional attachment. And Dr. Sid Salazar kept me focused and grounded in reality and provided realistic perspectives on the utility of my work.

My dear friends, Patricia Low, Mark Maggette, and Shelley Ferguson have been a constant source of love and inspiration. In addition to being a true friend in all respects, Dr. Anthony Odozi's intelligence, compassion, and thought-provoking analysis has fostered my own deep reflection on the meaning of my work. Dr. David Facer's genuine interest has been an honor that I cherish beyond measure. My doctoral peers Roxanne

Kymaani, Taylor Peyton-Roberts, and Emily Marx, have empowered me and cheered me all the way to the finish line.

I appreciate Caren Holtzman and Rusty Bresser who showed me the true value of a professional learning community and Cheryl Ritter, Danan McNamara, Kim Douillard, and Stacy Goldblatt who kept the torch of writing burning brightly in my life. Brad Callahan has been both a mentor and friend who provided unceasing guidance and support and Ryan Brock's conceptual and technical assistance was second to none.

Finally, I thank my parents, Colleen and Arnold Yarchever, who taught me the value of working hard and never to settle for less than what is possible. My children, Edward and Katherine Gallagher allow me to continually reflect on the purpose, power, and potential of education in people's lives.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	vii
TABLE OF CONTENTS.....	ix
LIST OF TABLES.....	xv
LIST OF FIGURES	xvii
CHAPTER ONE: INTRODUCTION.....	1
Statement of the Problem.....	4
Purpose.....	5
Research Questions.....	6
Limitations of the Study.....	6
Significance of the Study	9
CHAPTER TWO: REVIEW OF THE LITERATURE	10
Definition of Key Terms.....	11
Teacher Quality.....	11
Teacher Effectiveness	12
Teacher Evaluation	13
Student Engagement	16
Instructional Observation.....	18
Instructional Effectiveness.....	19
Defining Instructional Quality	19
Traditional vs. Progressive Views	20
The Science of Teaching.....	22

The Instructional Task	26
The Instructional Core	26
Quality Indicators.....	28
A Working Definition of Instructional Quality	34
Measuring Instructional Quality	35
A Common Vision for Effective Instruction	37
Content-Area Expertise of Evaluators	41
The Importance of Measurement, Observation, and Feedback	42
Instruction Is a Significant Factor for Student Achievement.....	44
Link Professional Development to Specific Teacher Needs.....	46
Additional Findings	48
Improving Instructional Quality	49
A Theoretical Framework for Instructional Quality and Measurement.....	54
Content Knowledge, Pedagogy, and Classroom Culture.....	55
Rigor, Relevance, and Relationships	55
Participation, Critical Thinking, and Academic Language.....	57
Instructional Improvement.....	60
CHAPTER THREE: METHODOLOGY	62
Research Questions.....	62
Research Context	63
Research Design.....	64
Sample.....	64
The Dependent Variable	65

Preliminary Analysis of Missing Data.....	67
Data Collection	72
Instrumentation	72
Reliability and Validity.....	75
Data Analysis.....	76
Model Specification.....	78
Question 1(a): How much do Student Growth Trajectories Vary?.....	78
Model 1: Baseline Model: Null Model/No Predictors.....	78
Question 1(b): How Do Student Characteristics Affect Growth Trajectories?	80
Model 2: Adding Student Characteristics.....	80
Model 3: Adding a Random Slope for Change at the Classroom Level.....	81
Question 1(c): Do Classroom and Teacher Characteristics Affect Student Growth Trajectories?.....	82
Model 4: Adding Classroom-Level Predictors.	82
Model 5: Adding Teachers as a Fourth Level in the Hierarchy.....	83
Question 2: What instructional practices yield the highest and lowest levels of instructional quality?.....	84
Summary of Methodology	86
CHAPTER FOUR: RESULTS	86
Mathematics.....	87
Model 1: Null Model/No Predictors	87
Model 2: Adding Student Characteristics.....	88
Model 3: Adding a Random Slope for Change at the Classroom Level.....	91

Model 4: Adding Classroom-Level Characteristics.....	93
English/Language Arts.....	98
Model 1: Null Model/No Predictors	98
Model 2: Adding Student Characteristics.....	100
Model 3: Adding a Random Slope for Change at the Classroom Level.....	101
Model 4: Adding Classroom-Level Characteristics.....	103
A Summary of the Quantitative Study.....	107
Cross-Case Narrative Analysis of Instructional Practices	108
Immediate Engagement	111
High Quality: High Instruction/High Achievement.....	112
Low Quality: Low Instruction/Low Achievement.....	113
Scaffolded Academic Dialogue	115
High Quality: High Instruction/High Achievement.....	115
Low Quality: Low Instruction/Low Achievement.....	118
Real-world Connections.....	119
High Quality: High Instruction/High Achievement.....	119
Low Quality: Low Instruction/Low Achievement.....	123
Front Loading.....	125
High Quality: High Instruction/High Achievement.....	125
Low Quality: Low Instruction/Low Achievement.....	126
Differentiated Instruction.....	128
High Quality: High Instruction/High Achievement.....	128
Low Quality: Low Instruction/Low Achievement.....	130

Feedback/Conferring.....	131
High Quality: High Instruction/High Achievement.....	131
Low Quality: Low Instruction/Low Achievement.....	133
Structured Reflection	135
High Quality: High Instruction/High Achievement.....	135
Low Quality: Low Instruction/Low Achievement.....	138
Lesson Closure and Connected Homework	139
High Quality: High Instruction/High Achievement.....	139
Low Quality: Low Instruction/Low Achievement.....	141
Integrated Strategies.....	142
Ineffective Practices.....	144
Students Waiting–No Expectations	144
Copying.....	145
Public Reprimands for Off-Task Behavior	145
Calling on Students One at a Time	147
Rapid-Fire Questioning.....	148
Teachers Answering Their Own Questions	149
Distribution of Quality by School.....	151
Distribution of PAQT Factors.....	152
A Summary of the Qualitative Study	153
CHAPTER FIVE: IMPLICATIONS	154
Redefining Instructional Evaluation	157
Adopt a Unified Vision of Instructional Effectiveness.....	159

Fundamental Principles of Instructional Evaluation.....	160
Criterion-Referenced Observations of Instruction.....	161
Content-Focused Observations of Instruction.	163
Unannounced Observations of Instruction.....	164
Consistent Observations of Instruction.....	164
Differentiated Professional Development.....	167
Quadrant-Specific Support and Intervention	168
Domain-Specific Support and Interventions.....	173
Curriculum	175
Communities of Practice.....	176
Professional Efficacy	178
Educational Equity	184
Distribute Quality Teachers Equitably Across Schools Based on Student Needs..	185
Correct Ineffective Practices.....	187
Make Cultural Proficiency Mandatory	188
Conclusion	190
References.....	192

List of Tables

1. Estimates examining the effect of missing data on 2012 math outcomes.....	68
2. Estimates examining the effect of missing data on 2012 English outcomes.....	69
3. Estimates of math fixed effects for missing data.....	70
4. Estimates of English fixed effects for missing data.....	71
5. Standard deviations for parameters converted to z-scores.....	84
6. Estimates of mathematics fixed effects for model 1.....	87
7. Estimates of mathematics covariance parameters for model 1.....	88
8. Estimates of mathematics fixed effects for model 2.....	89
9. Estimates of mathematics covariance parameters for model 2.....	90
10. Estimates of mathematics fixed effects for model 3.....	91
11. Estimates of mathematics covariance parameters for model 3.....	92
12. Estimates of mathematics fixed effects for model 4.....	94
13. Estimates of mathematics covariance parameters for model 4.....	95
14. Estimates of mathematics fixed effects for model 5.....	96
15. Estimates of mathematics covariance parameters for model 5.....	97
16. Estimates of English fixed effects for model 1.....	98
17. Estimates of English covariance parameters for model 1.....	99
18. Estimates of English fixed effects for model 2.....	100
19. Estimates of English covariance parameters for Model 2.....	101
20. Estimates of English fixed effects for Model 3.....	102
21. Estimates of English covariance parameters for model 3.....	102
22. Estimates of English fixed effects for model 4.....	103

23. Estimates of English covariance parameters for model 4.....	104
24. Estimates of English fixed effects for model 5.....	105
25. Estimates of English covariance parameters for Model 5.....	106

List of Figures

1. An interactive view of effective teaching.....	53
2. A theoretical framework for instructional quality and measurement.....	55
3. PAQT observation form.....	74
4. Instruction/achievement matrix: all classrooms.....	109
5. Instruction/achievement matrices for math and English.....	110
6. Distribution of instructional quality for School #1 and School #2.....	152
7. Percent of lesson segments scoring 0–4 for each domain.....	153
8. Criterion-referenced feedback summary.....	163
9. Ineffective Practices.....	169
10. High-quality instructional practices.....	172

CHAPTER ONE: INTRODUCTION

Competent, dedicated, and well-performing teachers are any school's most important resource. Teachers are the professionals most directly responsible for helping all students to learn and students benefit or suffer from the quality of the teaching they receive. Moreover, any society is at risk when its schools fail to educate its children and youth. So, clearly, effective teaching must be assured; and the teaching profession, school boards, school administrators, and school faculties must recognize that teacher evaluation is a key means of providing that assurance. (Shinkfield & Stufflebeam, 1995)

The No Child Left Behind Act (NCLB) of 2001 mandates a "quality teacher" in every classroom. Although state licensure, content area expertise, and years of experience are important for ensuring a basic level of professional competence, they are clearly not sufficient for evaluating the quality of instruction that students receive in the classroom on a daily basis (Berry, Daughtrey, & Wieder, 2010; Darling-Hammond & Sykes, 2003; Goe, Bell, Little, O., & National Comprehensive Center for Teacher Quality, 2008; Goldhaber, 2002; Hanushek, 2011; Heck, 2007; Hill & Ball, 2004; Sanders & Rivers, 1996; Wenglinsky, 2000). Many researchers and educational leaders working to improve teacher evaluation systems use value-added models to infer instructional quality from student test scores (Goe, 2007; McCaffrey, Lockwood, Koretz, & Hamilton, 2003). However, one problem with these methods is that standardized tests measure the acquisition of specific content standards, but were not originally designed to discern

variations in instruction or to sort out how different teachers contribute to student learning (Goe, 2007; Schacter & Thum, 2004). Additionally, teachers and policymakers are skeptical of using test scores to judge teacher effectiveness because they fear teachers who work in the most challenging and demanding instructional environments will be penalized because of factors affecting student learning that are not under their direct control (Shinkfield & Stufflebeam, 1995; Newton, Darling-Hammond, Haertel, & Thomas, 2010).

Of course, observation of instruction is considered the “gold standard” for assessing the quality of classroom instruction (Matsumura, Slater, Junker, Peterson, Boston, Steele, & National Center for Research on Evaluation, 2006; Waxman, Hillberg, & Tharp, 2004) and should be an integral part of any effective teacher evaluation system (Marshall, 2012a). McDonnell (1995) noted that, “some aspects of curricular practice simply cannot be measured without actually going into the classroom and observing the interactions between teachers and students” (p. 310). These aspects include the coherence of a teacher’s presentations, the questioning techniques used by the teacher, pedagogical decisions that elicit student participation in the learning process, grouping strategies that foster a productive use of instructional time, the pacing of lessons, the relative emphasis placed on different topics within a given lesson, and the teacher’s use of formative assessment to guide and promote classroom discourse.

In order to measure instructional quality, evaluators need observation protocols aligned with what is known about effective teaching practices (Junker, Matsumura, Crosson, Wolf, Levison, & Wiesberg, 2006; Waxman, et al., 2004). Focused observations of this sort can capture information about how curricula are presented to

students and the ways in which teachers maintain or degrade the potential cognitive demand of the content (Matsumura, Garnier, Slater, & Boston, 2008). Several studies have demonstrated methods for observing and evaluating instruction (Danielson, 1996; Danielson, 2007; Schacter & Thum, 2004), but few districts use them to effectively monitor instructional quality or to provide constructive feedback to teachers.

The lack of access to consistent observational data and reliable information about the quality of instruction teachers provide often manifests itself in a deep sense of disillusionment at many levels in the profession (Daley, Kim, & National Institute for Excellence in Teaching, 2010; Weisberg, Sexton, Mulhern, Keeling, Schunck, & Palcisco, 2009). For example, teachers are disillusioned with colleagues who are ineffective or stagnate in their practice and the supervisors responsible for helping them improve. Teachers are also disillusioned with districts when they fail to provide adequate resources and professional development (Newmann, King, & Youngs, 2000). Many in the profession are disillusioned with unions because the idea that seniority trumps the quality of a teacher's instruction is tough to accept at face value – despite the fact that a majority of people agree that the rights of teachers need to be protected. In addition, the public is disillusioned with the education profession for their failure to adequately measure teachers' classroom performance. Berry et al. (2010) highlighted this dissatisfaction when they noted, “To some degree, we might say that while our teaching effectiveness shortfalls may be large, our shortfalls in *measuring* teaching effectiveness and student achievement are even larger” (p. 4). Measuring teacher effectiveness, therefore, has moved into a position of prominence for educational researchers; however, controversy

surrounds both the purpose and methods for efficient, fair, and effective instructional evaluation.

Statement of the Problem

Streamlining the instructional evaluation process will help ensure a sufficient return on the investment that districts make in their teaching forces by enabling more students to achieve the skills and knowledge needed for active engagement in the communities of the 21st century. Although educational researchers have been successful at identifying the specific instructional characteristics that promote student learning (Bowen, 2003; Clark, Kirschner, & Sweller, 2012; Gardner, 1983; Hattie, 2009; Hidi & Renninger, 2006; Marzano, 2007; Mercer, 1995; Walqui, 2006; Waxman, And & Shwu-Yong, 1997), a significant number of educators continue to use strategies that fail to engage students in rigorous and relevant classroom experiences (Bruner, 1996; Hanushek, 2011; McDougal, Saunders, & Goldenberg, 2007; Schacter & Thum, 2004). Despite some recent advances in measurement, researchers have yet to develop tools that allow leaders to measure how a teacher's instruction impacts student learning in a manner that is both reliable, valid, and feasible (Goe & Stickler, 2008). The lack of understanding about the qualitative differences between teachers' instructional outcomes puts the profession at risk of failing to address serious issues of inequity in access to high-quality instruction (Barber & Mourshed, 2007).

We know that observation of instruction and the analysis of instructional tasks both provide good contexts for measuring instructional quality (Cuban, 2006; Matsumura, Garnier, Slater, & Boston, 2008), however, they are seldom used effectively by educational leaders to inform evaluations, make decisions about professional

development, or ensure schools are staffed with highly competent teachers (Kimball & Milanowski, 2009; Fink & Resnick, 2001; Weisberg et al., 2009). A systematic method for reducing high levels of subjectivity is needed, and models of educational effectiveness that allow for the accurate, efficient, and useful assessment of classroom instruction must be developed (Boser, 2011).

Purpose

The purpose of this study is to understand the impact of instructional quality on student achievement and to describe instructional quality in such a way that the data can be used to improve instructional practices. The research presented in this study supports and builds on a growing body of evidence that content area expertise, high levels of pedagogical competence, and the ability to create a classroom culture that fosters intellectual engagement are the critical attributes of teachers who deliver high-quality instruction (Ball & Rowan, 2004; Cuban, 2006; Elmore, 1996; Nystrand, 1997). The theory under investigation is that the observable student variables of participation, critical thinking, and use of academic language can serve as proxies for critical teaching skills and can be quantified to describe differences in the ways teachers engage students in learning. Analyzing this kind of classroom instructional data at school and district levels enables educational leaders to more effectively monitor instructional improvement and promotes more effective professional development decisions (Fink & Resnick, 2001; Newmann et al., 2000). It also empowers leaders to better allocate resources in ways that transform teaching and learning and increase student academic achievement. Creating a significant change in the quality of instruction experienced by students in the classroom is the most important outcome of this study.

Research Questions

In light of the above-stated purpose, the following research questions were developed in an effort to address this problem. The first question frames the quantitative component of the study and is divided into three sub-questions, which match the analytical steps needed when using multiple levels of data. The second question guides the qualitative inquiry.

1. What is the impact of instructional quality on student achievement?
 - a. How much do classrooms vary in the average level and shape of student growth trajectories in mathematics and English/language arts over the course of a school year?
 - b. How much of the difference in student growth trajectories is accounted for by student factors beyond the teacher's control?
 - c. Is the variance between classrooms and teachers in levels of participation, critical thinking, and students' use of academic language associated with the average level and shape of student growth trajectories?
2. What instructional practices result in the highest and lowest levels of instructional quality?

Limitations of the Study

This study has several limitations that impact the generalizability of the findings. These include the number of observations conducted in classrooms, the number of schools represented in the study, observer bias, and limitations inherent in the instrument itself.

Because each classroom was observed only one time, it is possible that the measure does not reflect the true level of instructional quality that students experience on a daily basis. Even though care was taken to ensure the instruction was representative of each teacher's daily practice, making multiple visits is essential for convincing teachers to trust the validity of the instrument (Marshall, 2012a) as well as making valid inferences about the quality of instruction over time (Matsumura et al., 2006).

Of course, the findings of this study apply to the specific schools that participated and can in no way be generalized to the larger profession. A larger sample of schools would undoubtedly yield a wider range of instructional quality and a more substantial list of quality indicators and ineffective practices.

There were three levels of observer bias in this study. Because the researcher created the instrument, conducted all of the classrooms observations, and held a supervisory position in one of the schools included, the ability to ensure objective measures came into question. Including an additional school was an attempt to mitigate these concerns, however, the addition also presented additional challenges. At one school, participation was mandatory because the observations took place during the course of the researcher's regular duties as an administrator. At the other school, participation was dependent on agreed-consent. Also, the observation measures may have impacted by the position that the researcher held at each school.

Concerns about whether the instrument created for this study was able to truly capture the nuances of teaching that make instruction an effective and dynamic experience for students is another limitation. As Shinkfield and Stufflebeam (1995) noted, "Any astute evaluator is fully aware of the fact that there is no such thing as

uniform teaching behaviors nor that learning results only from what is occurring in the lesson being observed” (p. 30). Although the variables in this study are presumed to represent instructional quality, it would be a mistake to think of any chosen set of classroom or school indicators as theoretically complete (Heck & Hallinger, 2009). As such, this instrument represents the beginning of a longer process designed to accurately capture the nuances of a construct that is exceedingly complex.

In regard to measurement, although generating a collective understanding of the qualities of powerful instruction is of critical importance, it is wise to be at least somewhat skeptical of mathematical representations of real-world processes, especially when people are personally impacted by the outcome. Figuring out methods for representing and communicating about instructional quality requires that stakeholders be cognizant of both the risks and the benefits of experimenting with novel measurement models (Heck & Moriyama, 2010).

A primary risk involving this instrument is specification error, or the oversimplification of the instructional process for the purposes of measurement. For example, alignment of instruction with the grade-level content standards on which students are being tested would be an important consideration in terms of instructional quality. The instrument presumes to capture this variable within the critical thinking component of the theoretical framework, but the instrument does not directly evaluate the lesson’s alignment with content standards. Indeed, there may be many missing variables that would be important to this model, so it would be naive to believe that the instrument in its current form is theoretically capable of detecting all of the factors associated with instructional quality.

An additional limitation has to do with a lack of alignment between the concept of instructional quality and the methods used for measuring student achievement. While the instructional component emphasizes critical thinking and the use of academic language, the California Standards Test (CST) requires only one-word answers or phrases which are provided for the student in a multiple choice format. An ideal test of instructional quality would compare the performance of students on a common assessment that included expectations for critical thinking and extended writing. As the profession moves to new student assessment systems based on the Common Core State Standards, which emphasize higher-order thinking and extended student constructed responses, there is potential for the instrument to demonstrate even stronger effects.

Finally, for highly skilled teachers, the instrument may not be able to capture the complex nuances that make high-quality instruction a powerful experience for students. There is a threshold at which expert instruction is beyond the scope of this instrument to measure. For example, when teachers differentiate instruction, teach small groups, or facilitate collaborative group work that is highly individualized, it is no longer possible to render class scores for critical thinking and language since the score varies based on each individual's involvement with his or her task.

Significance of the Study

Despite these limitations, there are some important reasons to move forward in the quest to better understand the concept of instructional quality. Accurately measuring the contributions that effective teachers make to children's learning is important not only to understand the characteristics and attributes of high-quality teaching, but also to assist teachers who are having difficulty (Weisberg et al., 2009). Identifying and supporting

teachers at all levels of the quality continuum cannot happen without effective measurement practices. Even if a measurement approach is at first inadequate, adopting such an approach has the potential to improve communication, understanding, consensus, and even arguments about the qualities of effective teaching practices; in turn, these discussions have the potential to produce improved measures (Glaser, 1976). The current variability in teacher effectiveness between classrooms results in significant educational advantages for some students who have access to high-quality instruction and disadvantages for those who do not (Aaronson, Barrow, & Sander, 2003; Heck & Moriyama, 2010; Ladwig, 2007). As a result, many students are denied opportunities to learn the skills and concepts needed for success in school. Understanding how individual teachers enhance or inhibit student learning based on their teaching practices increases the profession's ability to utilize instructional capacity to produce positive change.

CHAPTER TWO: REVIEW OF THE LITERATURE

This review of literature related to instructional quality was conducted in several phases. In the first subsection, I clarify the meaning of key phrases used by researchers and policymakers in discussions concerning the teacher evaluation process. These include: teacher quality, teacher effectiveness, teacher evaluation, student engagement, instructional observation, and instructional effectiveness. It is important to note that none of these constructs in their current form are being used effectively to assess classroom instructional practices on a wide scale. Second, I describe the evolution of the construct of instructional quality, including its history and development as an indicator of expert teaching and a valid predictor of student learning. Third, I review several recent studies

that demonstrate successful measurement of the construct and lessons learned for facilitating the instructional improvement process. Finally, I close with a theoretical framework for observing, measuring, and improving instructional quality on a wide scale.

Definition of Key Terms

Teacher Quality

Most national, state, and local departments of education define “teacher quality” or “highly qualified teachers” by looking at easy-to-measure teacher characteristics like post-baccalaureate coursework, subject-matter education, professional degrees, professional exams, years of experience, specialized certification, and evidence of participation in continued learning and professional development (Goe & Stickler, 2008). It is important to point out, however, that although these factors are important considerations for entry into the profession, only two—content area expertise in mathematics (Hill, Rowan, & Ball, 2005) and years of experience during a teacher’s first five years—have been shown to have any direct or significant impact on student achievement or student learning (Berry et al., 2010; Cohen, Raudenbush, & Ball, 2003; Darling-Hammond, 2000; Darling-Hammond & Sykes, 2003; Goldhaber, 2002; Hanushek, 2011; Heck, 2007; Sanders & Rivers, 1996; Wenglinsky, 2000). Importantly, Aaronson et al. (2007) found that 99% of the total variation in instructional quality is unexplained by teacher certification or advanced degrees and concluded that characteristics that are not easily observable in administrative data are driving much of the dispersion in instructional quality.

Teacher Effectiveness

The onset and use of sophisticated analytical tools that allow researchers to link student performance outcomes with specific teachers and schools has dramatically changed the way the profession conceptualizes teacher effectiveness. In this new era, teacher effectiveness most often refers to value-added measures that attempt to measure the contribution that each teacher makes to his or her students' learning, as indicated by higher-than-predicted increases in student achievement test scores (Goe & Stickler, 2008). These measures involve multilevel modeling techniques that are used to infer quality from residual estimates of student test scores after controlling for student demographic characteristics (Goe, 2007; Heck, 2009; McCaffrey et al., 2003). As was the case with teacher quality, studies that examine the use of value-added measures consistently indicate that the majority of variation in teacher effectiveness is due to "unobserved" variables.

This highlights a critical problem with using methods that depend on standardized tests to measure teacher effectiveness. Although the tests are meant to assess the acquisition of specific content standards, they were not originally developed to discern variations in instruction or to sort out teacher contributions to student learning (Goe, 2007). Teachers and policymakers are skeptical of these inferential measures because, although student progress is linked to individual teachers, they fear teachers who work in the most demanding instructional environments will be penalized because of factors affecting student learning that are not under the teacher's direct control (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012). Newton et al. (2010) studied value-added modeling as a measure of teacher effectiveness and found that using test

scores alone can both overestimate and underestimate a teacher's effectiveness based on the particular course, grade, or students the teacher is assigned. Not surprisingly, Newton et al. (2010) recommended that future research should "develop strategies for taking into account the various factors that may influence student achievement gains, so that the effects of teachers on student learning can be properly understood" (p. 20).

Teacher Evaluation

The most fundamental purpose of an evaluation is to *improve* both the individual's and the institution's performance (California Teachers Association, 2012; Danielson & McGreal, 2000; Patton, 2008; Stronge, 2006). The primary method for assessing instruction in the classroom is through teacher performance evaluations. Although the practice is common in almost all districts across the country and throughout the world (Sullivan, 2001), the definition of *attributes*, the methods for assessing quality, and the level of support provided when improvement is needed vary widely (Gordon, Kane, & Staiger, 2006; Kennedy, 2008; Kyriakides & Campbell, 2003). Ineffective teaching often goes unaddressed because tools such as professional teaching standards and evaluation instruments are not used to effectively communicate a common vision for high-quality instruction (Jacob & Lefgren, 2008; Kyriakides & Campbell, 2003; National Board for Professional Teaching Standards, 2002).

Moreover, due to union contracts, teachers in most states are given lifetime tenure after two or three years. This makes it difficult to terminate a teacher for ineffective performance (Kyriakides & Campbell, 2003; Unesco, 1997). As a result, even though schools and districts are held to high standards of accountability for student achievement based on test scores, teachers are not held accountable for high levels of teaching

(National Board for Professional Teaching Standards, 2002; Wang, 1998). This inability to address mediocre teaching practices through effective evaluation practices contributes directly to America's widening achievement gaps, since students with any kind of disadvantage desperately need effective teaching (Marshall, 2012a; Waxman, et al., 2004). In fact, there is no other socially significant profession where the employees are so insulated from accountability (Gordon, Kane, & Staiger, 2006).

Frase and Streshly (1994) studied six different school districts in the United States and found that, despite evidence of very poor instructional practice, none of the teachers were rated "below standard" on their annual performance evaluations. In New York City in 2008, three out of 30,000 tenured teachers were dismissed. The percentage of teachers dismissed for poor performance in Chicago between 2005 and 2008 was 0.1%. In Denver during the same time period, zero percent of teachers were let go for poor performance. Even though many teachers struggle to actively engage students in authentic learning, more than 99% of all teachers in the United States are rated "satisfactory" by their evaluators year after year (Weisberg et al., 2009). Berry et al. (2010) argued that "there are ineffective teachers in schools everywhere, but there might be fewer if our evaluation systems offered them constructive tools for improvement" (p. 4).

When effectively implemented, evaluation is the process of comparing an individual teacher's documented job performance with previously established roles, responsibilities, and research-based performance standards (National Board for Professional Teaching Standards, 2002; Stronge, 2006). An effective, ongoing, and systematic evaluation process identifies both strengths and areas of need and relates those

factors directly to the teacher's documented performance. The final evaluation captures both the merit and next steps of the teacher and also quantifies their performance in terms of student learning, but this is rarely a systematic or consistent process (Stronge, 2006). If evaluators differ substantially in the degree to which their ratings correlate with student achievement or if evaluations do not accurately reflect a teacher's skill in the classroom, teachers could receive accolades and/or consequences that are not justified or defensible (Kimball & Milanowski, 2009).

Most teacher evaluation instruments contain a list of components indicative of what it means to be an effective teacher. In his work on the Teacher Assessment Project, Collins (1990) noted five criteria for an effective teacher: (a) is committed to students and learning, (b) knows the subject matter, (c) is responsible for managing students, (d) can think systematically about his or her own practice, and (e) is a member of the learning community. Most teacher evaluation instruments used in districts contain similar criteria, but methods for assessing proficiency on each criterion vary greatly depending on who is evaluating and for what purpose (Jacob & Lefgren, 2012). Although Collins' list is seemingly comprehensive, it is inherently subjective; for example, how does one measure a person's ability to think systematically about one's own practice? Correctly evaluating each of these components is dependent on each evaluator's ability to correctly interpret and measure agreed-upon standards of performance (Eisner, 2004).

Every teacher possesses qualities that promote learning, and every teacher has areas of practice that can be improved. The problem is that teachers are in different places in terms of their ability to teach well, and defining consistent expectations for teachers while at the same time addressing individual strengths and weaknesses is a

complex process that is not easily distilled into a checklist or rubric (Cuban, 2013).

Because there is no agreed-upon standardized measurement, the marks teachers receive for their performance depend upon the background knowledge and experience of the evaluator as well as the evaluator's own understanding, or lack thereof, of what effective teaching is (Eisner, 2004). This lack of consistency and coherency undermines the improvement process and results in a lower level of trust among teachers that their evaluations will be valid (National Board Resource Center, 2010).

Student Engagement

Student engagement is a broad term that traditionally encompasses behavioral, emotional, and cognitive dimensions (Fredricks, Blumenfeld, & Paris, 2004). Student engagement in academic instruction is a significant factor associated with student achievement that results from students finding personal meaning in their learning and can serve as a critical observable outcome for measuring instructional effectiveness (Dooner, Mandzuk, Obendoerfer, Babiuk, Cerqueira-Vassallo, Force, & Roy, 2010; Harris, 2008). Too often, students sit in classrooms disconnected from their teachers, uninterested in the topics they are expected to study, and distracted by the forces in their environment that propel their attention away from intellectual endeavors (Cuban, 2013; Lounsbury & Clark, 1990). Scenarios like these are especially pronounced in schools serving poor and primarily Black and Hispanic students, where dropout rates are more than double those of schools consisting of White and Asian students (California Department of Education, 2012). Education reforms have attempted to close gaps in achievement by increasing student engagement in learning, but despite legal mandates (NCLB, 2001; IDEA, 2004), curricular innovations, teacher training, a wealth of professional literature, advice, and

methods about best practices, the gaps stubbornly persist (Bowen, 2003; Ferguson, 2003; Hampel, 1996; Haycock & Crawford, 2008; Picower, 2009). In fact, the very persistence of the problem reminds us that increasing student learning and achievement will require a sophisticated set of interventions that have a long-term impact on the ways teachers engage students in the active construction of new knowledge (Assor, Kaplan, & Roth, 2002; Bransford, Brown, & Cockling, 1999).

In an effort to reduce the impact of insufficient instruction on student learning, many practitioners and researchers have defined and investigated methods for maximizing student engagement (Bowen, 2003; Elmore, 1996; Hancock & Betts, 2002; Harris, 2011; Marzano, 2007). Teachers who provide effective instruction are cognizant of research-based best practice as they work to provide access to meaningful, relevant learning. In her phenomenological inquiry into middle school teachers' perceptions of student engagement, Harris (2008) identified six qualitatively different conceptions or levels of engagement: (a) behaving, (b) enjoying, (c) being motivated, (d) thinking, (e) seeing purpose, and (f) owning. While the first three levels relate primarily to a student's willingness and desire to participate, the final three emphasize the cognitive dimension of student engagement (Harris, 2008). Cognitive engagement encompasses deep and creative thinking about the concepts students are studying. As such, it includes thinking about what one knows and does not know; using meta-cognitive practices to plan, monitor, and evaluate progress and effort; and rehearsing, summarizing, and elaborating on material in order to better organize and understand it (International Center for Leadership in Education, 2002; Linnenbrink & Pintrich, 2003; Pintrich & De Groot, 1990; Tyler & Boelter, 2008). Although researchers have shown that engagement in both

critical thinking and meta-cognitive practices are important factors for student learning, these concepts have not yet been integrated into mainstream instructional evaluation protocols (Bowen, 2003; Schraw & Robinson, 2011).

Instructional Observation

Direct observations are considered the “gold standard” for assessing the quality of classroom instruction (Matsumura et al., 2006; Waxman, et al., 2004). Studies (Bowen, 2003; Matsumura et al., 2008) have shown that different teachers can use the same curricular materials with their students (as specified in a district’s scope and sequence plan, for example), but conduct discussions and engage students in assignment tasks that provide very different opportunities for students to deepen their comprehension and develop their academic skills. In her historical review of the development of the Opportunity to Learn (OTL) construct, McDonnell (1995) emphasized the value of observation as the best method for determining whether or not a teacher’s instruction is effective, further stating that curricular practice cannot be accurately measured without going into classrooms and observing the interactions between teachers and students. Observing student interactions illuminate the effects of discourse practices and grouping strategies on student participation, the degree of emphasis placed on core content within a given lesson, and the coherence of teachers’ presentations (Cazden, 2001).

The challenge comes when it is time to measure effectiveness on specific observational criteria. Many studies show that it is difficult for supervisors to use standards and rubrics effectively and consistently (Jacob & Lefgren, 2008; Kimball & Milanowski, 2009; McDougall et al., 2007; Schacter & Thum, 2004). In addition, the intensive time and resource demands associated with interpreting observational notes and

the cost of training observers so they understand the theoretical framework behind the rubrics represent significant system-level challenges. This may be the reason that many schools and districts do not use observation rubrics with the degree of certainty needed for teachers to take action on their recommendations (National Board Resource Center, 2010).

Instructional Effectiveness

Instructional effectiveness encompasses all of the distinct aspects of instruction that contribute to creating a change in the way students understand content (Ball & Rowan, 2004; Darling-Hammond, 2007; Fink & Resnick, 2001). It is different from teacher effectiveness in that it pertains directly to the interactions that happen in the classroom between teachers and students and does not include other important aspects of being a teacher like professional conduct, fulfilling non-classroom responsibilities, and being involved in extra-curricular activities. Instructional effectiveness is apparent when substantial and observable evidence exists that student learning has occurred (Danielson, 2007).

Defining Instructional Quality

The concept of instructional quality in the United States has been a source of exhaustive debate ever since the expansion of compulsory education began in the 1850s. The sheer quantity of teachers needed at that time created a strain on the intellectual and pedagogical capacity of the teaching force that still exists today (Reginier, 1994). As more and more students from increasingly diverse backgrounds began to utilize the public school and the goal of schooling became to “create a system of schools that could provide minimal education and basic socialization for masses of previously uneducated

citizens” (Darling-Hammond, 1997), levels of teacher expertise were significantly diminished. Since then, a variety of political and social forces have impacted our schools as the purpose of schooling has evolved, and responsibility and accountability for outcomes have been debated (Chelimsky, 1998; Donato & Lazerson, 2000). It has been difficult as a profession, as well as a society, to address the basic philosophical question of how to ensure equality of opportunity in education (Darling-Hammond, 2010; Marshall, 2012a; Tyack & Cuban, 1995). Rather than addressing the deeper, more complex issues of effectively educating an entire population, traditionalists and progressives engaged in pedagogical wars that have been more recently undertaken by proponents and adversaries of standardized testing practices (Henig & Stone, 2008). These debates have significantly impacted instructional quality in all of the major content areas (Bertrand, 2003).

Traditional vs. Progressive Views

The traditionalist’s perspective advocated for an academic curriculum focused on developing students’ subject matter skills and competence (Ravitch, 2000). Over time, as teachers became more and more dependent on commercially produced resources, traditionalists came to be associated with the direct instructional approaches and teaching methods prescribed in textbooks and teacher’s guides. These practices required students to recall information, recite factual knowledge, and use contrived texts, basal readers, and workbooks (Cuban, 2006; Pearson, 2004). Teachers were encouraged to use scripts from textbooks to teach subjects rather than operating from a position of deep knowledge about content. The teacher was seen primarily as responsible for the transmission of the prescribed curriculum, rather than as an expert in his or her field (Ladwig, 2009; Reginier,

1994). Direct instructional approaches came to be associated with formulaic practices and the “pedagogy of poverty” (Freire, 1970) which stimulated reformists to call for more interactive teaching methods that engaged students in critical thinking and reasoning within relevant and applicable contexts (Hampel, 1996; Schraw & Robinson, 2011).

The original goal of progressive education, characterized by the work of John Dewey (1910; 1916; 1934), was to make an “explicit attempt to change the core of schooling from a teacher-centered, fact-centered, recitation-based pedagogy to a pedagogy based on an understanding of children’s thought processes and their capacities to learn and use ideas in the context of real-life problems” (as cited in Elmore, 1996, p. 7). As student needs became the focus of instructional decision-making, the progressive movement came to be associated with discovery learning, whole language, new math, and differentiated instruction (Hargreaves, 2012; Pearson, 2004). Instruction based on the ability and interests of students, rather than the cognitive demand of the content, resulted in academic expectations being lowered for students who did not possess the prior knowledge needed to master more sophisticated content (Ravitch, 2000). Newmann, Marks, and Gamoran (1996) lamented that the ineffective implementation of progressive approaches to instruction were “leading down an illusory path where student participation in activities became an end in itself, regardless of the intellectual quality of students’ work” (p. 281).

The publication of *A Nation at Risk* (National Commission on Excellence in Education, 1983) reinforced the legitimacy of both Dewey’s original ideals and the traditional view that consistent, high expectations result in the maximum amount of learning. It accused the profession of failing to address the most common and

problematic form of instruction, which was “emotionally flat, intellectually undemanding, and un-engaging” (Elmore, 1996). *A Nation at Risk* (1983) stated that, “the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a nation and a people” (National Commission on Excellence in Education, 1983). The document revealed that instructional quality was a significant area of concern because of notable inequities in students’ opportunities to learn and the deterioration of America’s international standing in education.

The Science of Teaching

The verdict described in *A Nation at Risk* (1983) harkened back to the work of John Carroll (1963) who 20 years earlier showed that students vary in the amount of time it takes them to acquire academic competence. Carroll (1963) demonstrated that variations in the quality of instruction greatly influenced the differences in time required for students to learn. He developed a scientific model for evaluating the quality of instruction based on three criteria: (a) learners must be clearly told what they are to learn; (b) they must be put into adequate contact with learning materials; and (c) the steps in learning must be carefully planned and ordered. These prescriptive criteria led to the development of subject area content standards, mandates for access to standards-based curricula (McLaughlin, Shepard, & National Academy of Education, 1995), and a focus on pedagogy as a critical component of the instructional process (Shulman, 1986).

Carroll’s (1963) research demonstrated that a learner’s ability to understand instruction was dependent on the student variables of language comprehension and perseverance. Carroll described perseverance as the amount of time a student needed to learn a given task to an acceptable criterion of mastery under optimal conditions of

instruction. The combination of perseverance with student motivation was termed “aptitude.” High aptitude was indicated when a student needed a relatively small amount of time to learn, while low aptitude was indicated when a student needed much more than average time to learn.

The OTL construct, which grew out of Carroll’s work, was defined as the amount of time allowed for learning. It implied that, unless all students have the same aptitude, the teacher would be responsible for providing individual students the requisite time to learn (McDonnell, 1995). Two other OTL variables related to student achievement were the quality of instruction and the ability to understand instruction. If the quality of instruction was less than optimal, then more time was needed for learning; similarly, if a student was lacking in the ability to understand the instruction, more time was also needed. This implied that the teacher had a significant role to play in the development of student aptitude. Bruner (1966) suggested that, in this instance, it would be important for the teacher to develop methods to assess student understanding so that the appropriate levels of instructional scaffolding could be provided.

In terms of defining quality, although Carroll’s (1963) model was admittedly simplistic, it emphasized the inherent complexities of applying psychological principles to the design of instruction (Glaser, 1976; Haertel, Illinois Univ., & And, 1980). Carroll’s (1963) evidence suggested that when the quality of instruction and opportunity to learn were properly managed through effective teaching practices, student perseverance would take care of itself. This implication made it clear that student motivation was significantly influenced by the quality of instruction and the relationships that developed between the teacher and students both individually and collectively

(Bruner, 1966). Carroll maintained that even an oversimplified model was better than no model at all when addressing complex phenomena like instruction.

Carroll's model was later taken up by Benjamin Bloom (1968), who postulated that 95% of students could come to achieve mastery of standard school subjects if attention were paid to increasing the ratio of time spent to time needed, either by increasing the time spent learning (the numerator of the ratio) or by reducing the time needed to learn (the denominator), or both. His Mastery Learning model demonstrated that improving the quality of instruction and enhancing students' motivation and aptitudes were the critical variables for maximizing learning. He showed that when students receive feedback and are given corrective, individualized help, they overcome the obstacles that would otherwise cause them to fall behind (Bloom, 1978). He also showed that an environment that focused on each student's individual needs as opposed to an environment that treated students identically became much more cooperative and collaborative, while the traditional setting became increasingly competitive and caused fewer students to reach high levels of mastery. These ideas have continued to evolve in the profession and currently reside in the response to instruction and intervention (RtI²) model. The RtI² model was designed to reduce the inappropriate and disproportional placement of struggling learners in special education programs and to ensure that instruction is sufficiently responsive to the needs of all students including culturally and linguistically diverse youth (Hernández Finch, 2012).

In Carroll's (1989) 25-year retrospective on his model and the evolving concept of OTL, he reiterated his original cautionary assertion that "time as such is not what counts, but what happens during that time" (p. 27). Surprisingly, of all the variables in

Carroll's (1989) studies of educational effects, only quality of instruction failed to show a consistently significant effect. He reasoned that this was probably due to problems in measuring the variable and closed his retrospective with the following assertion:

For all students, quality of instruction must be maintained at the highest levels appropriate in a given case. Teachers must be not only intelligent and competent as classroom managers but also adequately knowledgeable about the subject matter they teach. Instructional materials should be prepared and sequenced on the basis of the best research on the cognitive skills involved, and matched in a challenging way to students' levels of aptitude, skill, and knowledge. Instruction should clearly specify what is to be learned. Such procedures of mastery learning as formative testing, corrective feedback, and so forth should be used whenever they are appropriate and feasible. (p. 30)

It is clear from this quote that Carroll associated instructional quality with teacher competence and ability, emphasizing the importance of classroom management and a solid understanding of subject matter. He recognized the importance of research-based curriculum materials and implied that even students with lower aptitude could benefit from challenging assignments. His model demonstrated very high expectations for teaching, and he made it clear that ongoing assessment and feedback were critical to effective student–teacher interactions.

The Instructional Task

Glaser (1976) suggested that the science of instructional psychology, in other words, the study of the process of instructional design that results in effective student learning, should guide teachers' decision-making in the formulation and implementation of instructional tasks. His instructional processes framework emphasized the concept of pedagogy as a linking structure that could be accumulated into a body of professional practices. Pedagogy would serve as the link between psychological theories of learning and cognition and teaching as an applied science in much the same way that doctors translate the biological sciences and engineers translate the physical sciences into real-world activities that impact human lives. The primary role of the teacher, therefore, is to create instructional tasks that integrate the interactive components of effective teaching (Bowen, 2003).

The Instructional Core

Elmore (1996) referred to this active, facilitated learning process as the instructional core of teaching. For Elmore, the instructional core represents "a teacher's understanding of the nature of knowledge and the student's role in learning, and how these ideas about knowledge and learning are manifested in teaching and class work" (p. 2). He also emphasized the central role that teachers play in the classroom because they are in positions of authority and, as such, make important decisions about what and how students learn.

Similarly, Cohen and Ball (1999) conceived the instructional core as a way to represent the critical space in the learning process where authentic change in understanding takes place. They postulated that "instruction consists of interactions

among teachers and students around content” (Cohen, Raudenbush, & Ball, 2003, p. 122). Hiebert and Grouws (2007) added an evidentiary component to the definition, saying that “teaching consists of classroom interactions among teachers and students around content directed toward facilitating students’ achievement of learning goals” (p. 372). Darling-Hammond (2011) re-emphasized the role that prior knowledge plays, stating that at the very core of teaching is the task of helping students to make connections between what they already understand and the new concepts, information, or skills we want them to learn. In this context, learning requires that students engage in active ongoing re-construction of their conceptual understanding (Darling-Hammond, 2011).

Darling-Hammond (2011) also emphasized the importance of assessing student learning not just prior to or after instruction, but continuously as lessons and units progress. A teacher’s ability to adapt tasks to address student needs, both individually and collectively, represents a very high level of instructional capacity (Bowen, 2003). The ability to engage in thoughtful, ongoing assessment during instruction is dependent on clear standards, constant feedback, effective scaffolds and supports, and tasks that include built-in opportunities for students to revise their work (Bowen, 2003). In this teaching paradigm, teachers develop and effectively manage a collaborative classroom in which all students have membership (Darling-Hammond, 2011). Facilitation of learning on this level highlights the inter-related nature of standards, curriculum, assessment, and the teacher actions responsible for bringing them all together to create a cogent, engaging experience for students.

Quality Indicators

Wang, Haertel, and Walberg (1990) conducted a comprehensive review of all peer-reviewed literature and significant book chapters related to how children learn in an attempt to discern the critical variables associated with student achievement in school. Their intent was “to provide a synoptic view of the entire panoply of variables” (p. 31) associated with student learning outcomes. Their synthesis resulted in 228 items organized into six general categories. The last category, which was directly concerned with the quality of instruction in the classroom, included the following teacher behaviors: (a) establishing efficient classroom routines and communicating rules and procedures; (b) using clear and organized direct instruction; (c) maximizing time on task (i.e., amount of time students are actively engaged in learning); (d) using assessment as a frequent, integral component of instruction; (e) using questioning/ recitation strategies that maintain active participation by all students (i.e., group alerting); (f) creating an environment where students respond positively to questions from other students and the teacher; (g) frequently calling for extended, substantive oral and written response (i.e., not one-word answers); and (h) establishing cohesiveness among students (i.e., developing an environment where members of the class are friends sharing common interests and values and emphasizing cooperative goals).

Wang et al.’s (1990) emphasis on teacher behaviors demonstrated the important impact that teacher decisions have on student learning. “The highest rated instructional variables suggested that the key to effective instructional design was the flexible and appropriate use of a variety of instructional strategies, while maintaining an orderly classroom environment” (p. 35). According to their study, the items most important to

increased learning outcomes were those that were directly tied to students' engagement with the material to be learned. This focus on instruction was in direct contrast to the Coleman report (1966), which declared home and student factors to be the most significant predictors of student achievement. Wang et al. (1990) suggested that, from kindergarten through grade 12, across a range of content areas and educational contexts, quality and quantity of instruction are roughly equal to the importance of student characteristics and out-of-school contextual items. Their study concluded that "classroom management, climate, and student-teacher interactions represented an important constellation of variables related to effective instruction" (p. 37) and that the teacher's role is a critical factor in student learning.

Growing concern about all children having equitable access to quality instruction led Porter (1991) to underscore the right and the responsibility of the American public to know the nature of the education being provided to children. He insisted that "information be made available to describe the probability that a particular type of student will receive good teaching of worthwhile content" (p. 13), and that this could not happen unless we began looking directly into classrooms. Porter began identifying school process indicators that included both organizational characteristics and instructional characteristics.

In his model, instructional characteristics were subdivided into curriculum quality and teaching quality. Porter focused on the "enacted" as opposed to the "intended" curriculum because the enacted curriculum was more representative of what was actually taught in classrooms. He focused on instruction rather than student achievement because "schools have more direct control over what is taught than they have over what students

learn” (p. 25). In determining instructional quality indicators, Porter (1991) emphasized the importance of indicators having enduring value and conceptual clarity, both in intention and in technical definition. He said they should be few in number and they should reflect a central feature of the education system. In order to ensure utility, Porter maintained that indicators must be easily understood by a broad audience, so that they effectively serve the information needs of policymakers and the public.

Porter showed that quality instruction is not about a checklist of attributes; instead, he emphasized the importance of the interactions that happen as a result of a collection of important variables. “To reflect the complexity of instructional quality while maintaining the need for parsimony, an indicator may be a function of several separate characteristics” (Porter, 1991: p. 24). These characteristics included the extent to which teachers hold goals consistent with desired student outcomes. Teachers who have different goals for students will have different instructional expectations and outcomes. The extent to which teachers accept responsibility for student success or failure in reaching desired student outcomes is a second indicator of quality because how teachers respond when students don’t learn determines whether additional instruction takes place or not and how interventions are crafted. A third indicator is the extent to which teachers are clear to their students about what is to be learned and why. How teachers frame the learning for students impacts students’ engagement and their motivation to understand the content in increasingly greater depth. Porter’s contribution was consistent with a socio-cultural theory of teaching and learning (Tharp & Gallimore, 1988; Vygotsky, 1986; Wertsch, 1991), which highlights the importance of students having opportunities to construct their own knowledge, rather than simply memorize or recite information. Porter emphasized

that the role of the teacher was to create tasks that fostered constructive interactions with content and that this was dependent on the teacher's ability to understand student misconceptions. If students struggled to engage with the material, he encouraged teachers to reconfigure tasks until the desired student outcomes were achieved. This responsibility for outcomes alluded specifically to the concept of teacher efficacy (Bandura, 1993; Guskey & Passaro, 1993) and would provide the basis for the formation of professional learning communities where teachers work collaboratively to solve the problems of teaching by understanding instruction and its connections to student learning (Bowen, 2003; Dufour et al., 2006).

Resnick and Hall (2001) investigated the connections between motivational and cognitive research and developed *The Principles of Learning* based on the convergence of theories emphasizing Carroll's idea that intelligence is not a fixed construct, but is learnable and teachable. The *Principles of Learning* maintained that academic rigor and a progressively deepening understanding of concepts takes place in a classroom environment that purposefully nurtures and supports intellectual collaboration as the norm. In their model, curriculum and assessment are organized around the mastery of major core concepts. Students complete challenging assignments, raise key questions, justify explanations, and reflect on their learning. Students are critical thinkers who synthesize, apply, and interpret concepts and construct solutions to problems based on those interpretations (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Rath, & Wittrock, 2001; Bloom, 1956, 1978; Nasstrom, 2009; Webb, 2005, 2007). Students use language to explain, clarify, and challenge ideas as well as provide evidence to substantiate claims and arguments. Standards for student work are clearly articulated;

students have access to multiple proficient models; and students self-evaluate, self-monitor, and receive ongoing feedback and support on their progress toward mastery.

As knowledge grew about what constitutes an effective learning environment, Ferguson (2003) became primarily concerned with closing the achievement gap for underserved populations of students by ensuring they had access to high-quality instruction. He described three foundational elements of teaching called the “instructional tripod,” which included content, pedagogy, and relationships (Ferguson, 2003). This powerful metaphor highlights each element as being so critical that if even one of the legs of the tripod is too weak, the whole instructional episode will collapse. He found that there was a higher probability that this kind of collapse would take place in our nation’s schools that serve the most vulnerable children. Indeed, as of 2007, there remained serious concern about the ability of teachers to implement pedagogical approaches that integrate each of these domains in such a way that instruction results in positive outcomes for student learning (Hiebert & Grouws, 2007). Hiebert and Grouws (2007) showed that three important elements critical to effective instructional outcomes are still absent from American classrooms. One is the expectation that students will construct their own ideas about the application of a concept. Another is that students will understand multiple representations of a concept. The final missing element is the expectation that students will use their own language to explain their understanding. Their research emphasized that authentic questions asked by the teacher from a source of deep knowing are what facilitate deep thinking about meaning. Palmer (1998) asserted that good teaching cannot be equated with technique, nor can teaching be scripted for mass production by practitioners. Palmer’s work inspires educators to have the courage

to believe in their own ability to impact students in very personal ways. For Palmer (1990) good teaching “comes from the integrity of the teacher, from his or her relation to the subject and students, from the capricious chemistry of it all” (p. 10).

The critical, inter-related components of effective teaching described above are also consistent with three of the four components of the Praxis III Classroom Performance Assessment Criteria (Dwyer, 1998). The Praxis III replaced the National Teacher’s Exam (NTE) as the primary method for certifying beginning teachers. Praxis III focuses on assessing competencies in (a) organizing content knowledge for student learning; (b) creating an environment for student learning; (c) teaching for student learning; and (d) teacher professionalism (Scriven, 1994; Shinkfield & Stufflebeam, 1995). A teacher’s ability to be reflective about and accountable for high levels of instructional quality constitutes the fourth competency, which is teacher professionalism. This competency is represented by the interactions between all three areas of expertise and is embodied in the instructional tasks the teacher facilitates (Bowen, 2003; Hatano & Osuro, 2003; Porter, 1991). Once teachers attain this certification however, they are seldom held accountable for maintaining the professional expertise the initial assessment requires and instead fall into patterns based on their school environments (Weisberg, et al, 2009).

Taken together, Porter’s (1991) indicators, Wang et al.’s (1990) research, Resnick and Hall’s (2001) *Principles of Learning*, and Ferguson’s (2008) *Instructional Tripod* substantiate the complex and interactive nature of the teaching process and reveal an important connection between instructional decision-making and quality (Cuban, 2013). For example, one of Porter’s quality indicators, “holding goals consistent with desired

student outcomes,” implies deep knowledge of content standards, the developmental nature of concept exploration, an understanding of each student’s relation to the content to be presented, methods for assessing student understanding, and the ability to differentiate instruction based on the individual needs of each student. It is these interactions between content, pedagogy, and classroom culture that represent the complex nature of the teaching process, rather than simple isolated descriptors that often fall short in their ability to capture the essence of quality teaching.

The common thread running through each of these models is that learning takes place in an environment that supports the intellectual advancement of each individual student (Boettcher, 2007; Palmer, 1990; Vygotsky, 1986). It is not enough for a teacher to be certified in his or her content area. The teacher must also be savvy and experienced enough in pedagogy to construct and present tasks that facilitate each student’s learning (Bowen, 2003). This facilitation requires the teacher’s higher-order thinking and meaningful enhancement of the prescribed curriculum so that students are able to connect with and strengthen their understanding of important academic concepts (Hill, Blunk, Charalambous, Lewis, Phelps, Sleep, & Ball, 2008).

A Working Definition of Instructional Quality

At this point in the review, it is clear that instructional quality is a complicated construct that transcends arguments about traditional values versus progressive approaches (Henig & Stone, 2008). Synthesizing the work of the aforementioned theorists has led to a working definition of instructional quality that is meant to facilitate a common understanding about what is meant when we say that all children deserve a high-quality education. For this study, instructional quality refers to the level at which a

teacher facilitates multi-directional interactions with the class that result in authentic cognitive engagement and increased subject matter competence (Cohen, Raudenbush, & Ball, 2003; Goe & Stickler, 2008; Matsumura et al, 2006; Tyler & Boelter, 2008; Westgate & Hughes, 1997). Authentic cognitive engagement means that students are interested in learning, are thinking deeply about their task, and are learning to articulate increasingly sophisticated understanding of concepts (Elmore, 2009). The definition emphasizes the multi-directional nature of the interactions as the teacher directs instruction toward facilitating each student's achievement and learning (Nystrand, 1997; Olson, 2003; Westgate & Hughes, 1997). Cognitive engagement is measured through direct observation in the classroom. Subject matter competence, on the other hand, is the observed outcome associated with increased student achievement as measured by a variety of formative and summative assessments. Quality is represented by the interaction of cognitive engagement and subject matter competence. Facilitating multi-directional communication about and connections with content represents the art of teaching that is so difficult to capture in traditional measures of instruction (Hiebert & Grouws, 2007; Westgate & Hughes, 1997). This study attempts to illuminate this challenge.

Measuring Instructional Quality

It is clear from these descriptions that researchers have successfully articulated a vision of what effective instruction looks like and sounds like, but whether or not the majority of teachers are able to provide such an environment for learning remains an urgent concern. For many years, researchers have tried to link the quality of a teacher's instructional practices with the academic achievement of students. The problem is that,

although measures of student achievement are relatively stable, measures of teacher practices are not. In this second subsection, I review promising new measurement constructs and their associated instruments so that lessons can be learned from their implementation. The research described in this subsection of the review will focus on synthesizing studies that meet two criteria. First, the studies either used or developed a systematic method for observing instruction in the classroom and measuring its quality. Second, the studies used standardized assessments to link instructional practices with growth in student achievement. Studies excluded from this review are those that used survey or interview methods to assess instructional quality or those that inferred instructional quality from teacher qualifications or student test scores. Although surveys, interviews, and improved teacher test scores are efficient, cost effective, and may contain some signal about instructional quality, their ability to serve as accurate representations of a teacher's practice has not been substantiated (Aguirre-Munoz et al., 2006; Matsumura et al., 2006).

Five studies stood out in the literature that met both of the above criteria. While each sought to assess instructional quality in different ways, they all made significant contributions to our understanding of measurement. The studies shed light on the variety of methods that have been created for observing instruction and also the challenges associated with using observation on a wide scale. The most important challenge arises in the time and resources that are required for fair, accurate, and system-wide evaluation. Despite measurement advances, this still remains a significant obstacle. Other issues that are addressed by the studies include inter-rater reliability, disruption to the natural flow of teaching and learning, and the amount of time or artifacts needed for obtaining effective

measures of instructional practice (Matsumura et al., 2008). Despite these obstacles, researchers are pushing the work of measurement to higher levels of effectiveness and, together, the studies have important lessons to teach about measuring instructional quality and correlating those measures with student performance. These lessons include: (a) establishing a common vision for effective instruction, (b) ensuring evaluators have high levels of content-area expertise, (c) acknowledging the importance of measurement, observation, and feedback for ensuring standards for excellence are being met, (d) recognizing that instruction is a significant factor for student achievement, and (e) making certain that professional development is linked to the specific needs of teachers.

A Common Vision for Effective Instruction

Effectively assessing instructional quality—the level at which a teacher facilitates multi-directional interactions with the class that result in authentic cognitive engagement and increased subject-matter competence—requires a precise description of the details of how teachers use curriculum, assessments, content knowledge, and their pedagogical skills to enhance student learning (Glaser, 1976; Hiebert & Grouws, 2007; Nystrand, 1997). Porter (1991) advised describing instruction “at a level of detail that is neither so fine-grained that every classroom or policy looks unique, nor so crude that all classrooms or all policies look the same” (p. 18). He emphasized that “there is no perfect solution to the degree-of-specificity dilemma and that any resolution will be somewhat arbitrary and, so, subject to argument” (Porter, 1991, p. 18). The bottom line is that developing an observation protocol aligned with best practice is an essential part of any investigation of the measurement of instructional quality (Barber & Mourshed, 2007; Waxman, et al., 2004).

The Framework for Teaching (FFT) is one of the most famous attempts to create both a vision and a measurement of instructional effectiveness (Danielson, 1996, 2007). The FFT has become the standard for observational evaluation of instruction and is comprehensive enough for use with teachers across a broad range of experience. It uses multiple sources of evidence with very specific assessment rubrics to measure effectiveness. It is divided into 22 components, which are clustered into four domains of teaching responsibility. For each component, there are detailed descriptions of each level of teaching performance: Unsatisfactory, Basic, Proficient, and Distinguished. For example, the descriptor for an unsatisfactory ranking for component (1a), “Demonstrating knowledge of content and pedagogy,” says,

In planning and practice, teacher makes content errors or does not correct errors made by students. Teacher’s plans and practice display little understanding of prerequisite relationships important to student learning of the content. Teacher displays little or no understanding of the range of pedagogical approaches suitable to student learning of the content (Danielson, 2007).

Whereas, for the distinguished ranking, the descriptor says,

Teacher displays extensive knowledge of the important concepts in the discipline and how these relate both to one another and to other disciplines. Teacher’s plans and practice reflect understanding of prerequisite relationships among topics and concepts and a link to necessary cognitive structures by students to ensure understanding. Teacher’s plans and

practice reflect familiarity with a wide range of effective pedagogical approaches in the discipline, anticipating student misconceptions (Danielson, 2007).

These descriptions succeed in solidifying a common language for describing instructional quality and help develop a shared understanding of the range of instructional practices that comprise effective teaching. The rubrics are the most commonly used tools to date for evaluating and providing feedback on classroom instruction, and are used in hundreds of schools and districts nationwide.

Many other observation protocols have been developed that also succeed in creating a common vision for instructional quality. Newmann et al. (1996) studied 23 successfully restructured schools at the elementary, middle, and high school levels in the areas of both mathematics and social studies and formulated standards for the intellectual quality of instruction based on the cognitive demand of student assignments and the observed interactions between teachers and students. Their vision of quality was called “authentic pedagogy,” where students engaged in the active construction of knowledge through disciplined inquiry to produce discourse, performances, or artifacts that had value beyond certifying success in school. Similarly, Schacter and Thum (2004) identified 12 key components of effective teaching and created rubrics to measure the level at which teachers were able to implement the strategies in the classroom. Their 12 teaching performance standards included: teacher content knowledge, lesson objectives, presentation, lesson structure and pacing, activities, feedback, questions, thinking, grouping students, motivating students, classroom environment, and teacher knowledge

of students. Similar to the FFT, Schacter and Thum (2004) provided a rubric that described a range of proficiency for each standard from ineffective to exemplary.

Another model for effective evaluation of instruction is the Instructional Quality Assessment (IQA) (Matsumura et al., 2006), which was developed out of the growing need for a technically sound tool that could effectively measure quality of instruction on a large scale, yet, unlike the FFT, be a minimum burden on teachers, schools, and districts in terms of time and cost. Matsumura et al. (2006) drew heavily on the research synthesis *How People Learn* (Bransford et al., 1999) and conceptualized their instrument around four of the *Principles of Learning* (Resnick & Hall, 2001). The observable standards of academic rigor, clear expectations, self-management of learning, and accountable talk were each presumed to characterize expert teaching. Recognizing that student work also played an important role in the evaluation of quality, a second component of the IQA included an analysis of instructional tasks assigned by the teacher. Their vision was that expert teachers (a) help students develop a clear understanding of what they should know and be able to do; (b) set learning goals and monitor progress; (c) ensure students produce quality work, showing evidence of understanding, not just recall; and (d) utilize assessment tasks that require students to exhibit higher-order thinking. Matsumura et al. (2006) developed their instrument in an effort to make those values a reality in classrooms.

Although the authors of the previous studies constructed their own models for evaluation, observation instruments do not always need to be created from scratch. Aguirre-Munoz et al. (2006) adapted the Sheltered Instruction Observational Protocol (Echevarria, Vogt, & Short, 2000) to investigate the variability of OTL in middle school

classrooms serving high percentages of English Learners. Specifically, they studied English Learners' access to academic language instruction by observing the teacher's ability to provide comprehensible input, build background knowledge, and scaffold metacognitive strategies. The researchers linked their measures to student performance on the Language Arts Performance Assignment (LAPA). Their vision of effective instruction enabled English Learners to actively participate in learning based on the teacher's theoretical knowledge about language structures, the teacher's effective use of instructional strategies, and the teacher's ability to scaffold instructional conversations, analyze student writing, and collaboratively develop lessons using functional linguistic concepts.

These studies provide examples for districts aiming to improve their instructional evaluation procedures. The studies emphasize the importance of clearly articulating an agreed-upon vision for what it means to provide effective instruction. Any research-based protocol should reflect the instructional goals of the district and should be clearly understood and evenly and fairly applied to all teachers to facilitate accurate measurement, teacher buy-in, and ongoing instructional improvement (Cuban, 2013; Goe et al., 2008).

Content-Area Expertise of Evaluators

Using experienced evaluators who also have content-area expertise ensures more accurate assessments of instructional quality (Matsumura et al., 2006). Newman et al. (1991) conducted their observations and scored assessment tasks using researchers and teachers, who both had knowledge and experience teaching the subject. Likewise, Schacter and Thum (2004) recommended that districts employ and train multiple

evaluators to assess teachers so that they are able to provide accurate and reliable measures of teacher performance in a consistent and timely manner. Providing well-organized and purposeful rater training sessions is necessary to ensure evaluators are clear about precisely what aspects of instruction are under examination and to calibrate observations to increase inter-rater reliability (Goe, 2008).

Eisner (2004) conceptualized evaluation as educational connoisseurship and criticism. He writes that evaluation is connoisseurship when one knows what one is looking for, can recognize its presence and absence, can offer reasoned rationale for one's judgment, and can recognize subtle but significant differences. Eisner (2004) noted, "Connoisseurs are people who come to know, and critics are people who can render what they come to know in a language that is accessible to others and that enables others to 're-see' the work, the performance, or the object at hand" (p. 198). The ability of evaluators to assist educators in "re-seeing" instruction enables reflection on teaching, which is an essential attribute of an effective teacher's professionalism (Boyd & Fales, 1983; Cuban, 2013; Danielson, 2007; Dufour et al., 2006; Schön, 1983).

The Importance of Measurement, Observation, and Feedback

Once evaluation protocols are in place, it will take some time for teachers to familiarize themselves with the new standards. A common finding across all studies was that there was significant variation in observed instruction and, generally, instruction fell far below the highest levels of the proposed standards for instructional quality, regardless of the construct used (Matsumura et al., 2006; Newman et al., 1991). Aguirre-Munoz et al.'s (2006) research demonstrated that asking teachers what they teach and observing what and how teachers teach provide very different measures of instructional quality and

reveal significant discrepancies in student access to learning opportunities. For example, in surveys, teachers in their study mentioned the use of procedural scaffolding, a strategy related to Vygotsky's (1978) gradual release of responsibility, as an important instructional practice for English Learners. However, observations revealed that teachers tended to move directly from whole-group guided instruction to independent work; additionally, if they did use scaffolding, they used it ineffectively in ways that essentially denied English Learners access to rigorous work and conversations. In fact, no teachers in their study demonstrated proficiency with academic language instructional strategies, and their results underscored the need for systematic examination of observable variables to monitor the quality of instruction in classrooms. As a result of poor instruction, opportunities for negotiation of meaning and access to the group-level intensive support that is so essential for language development were greatly reduced, and this reality would not have been discovered had the study relied only on survey methods. Observation was essential to understanding how the dynamics between teachers and students either inhibited or enhanced learning.

Across studies, the over-reliance on direct instruction as the primary teaching method has been shown to significantly impact student engagement and, subsequently, student achievement (Aguirre-Munoz et al., 2006; Matsumura et al., 2006; Newman et al., 1991; Schacter & Thum, 2004). Although the profession has become increasingly expert at developing standards—most recently, Common Core State Standards—the difficult task ahead involves assuring teachers are able to teach them effectively, especially for students attending low-performing schools. Schacter and Thum (2004) found that classrooms with higher concentrations of Hispanic, limited English proficient, or low-

performing students tended to be taught by teachers who obtained lower scores for instructional quality. This observation revealed that, even though we know that high-quality instruction improves learning, the students most in need of that learning typically did not have access to it because they tended to be assigned to teachers whose instruction did not meet the standards. Similarly, Matsumura et al.'s (2006) IQA ratings of observed instruction demonstrated a wide variation in instructional quality, with most classrooms demonstrating only basic quality. The authors noted:

Teachers frequently did not build on, extend student contributions, or press students to explain their reasoning or give evidence for their assertions, and this was reflected in brief, surface-level student responses. The expectations for learning that teachers communicated to students also did not focus on high-level demands in the majority of classrooms.

(Matsumura et al., 2006, p. 26)

Even so, in spite of considerable obstacles, there are teachers and schools who demonstrate considerable progress toward high standards of quality (Matsumura et al., 2006; Newman et al., 1991; Schacter & Thum, 2004). This progress suggests that establishing standards of instructional quality and providing regular sustained feedback to teachers on the alignment between the agreed-upon vision and their actual performance can be quite useful in helping the profession improve their capacity for instructional evaluation (Newman et al., 1991; Regnier, 1994).

Instruction Is a Significant Factor for Student Achievement

If teachers' instructional practices are not specifically linked to gains in students' academic competence using standardized measures, the information gleaned from them

will be of little generalizable value in terms of improving instruction on a wide scale or informing any policy recommendations based on the findings (Aguirre-Munoz et al., 2006; Cuban, 2006; Hiebert & Grouws, 2007). By using a common, integrated set of standards to examine instruction, the studies in this review add significant empirical knowledge to the question of an important relationship between instruction and student achievement.

Kane, Taylor, Tyler, and Wooten (2011) found that teachers in the top value-added student achievement quartiles consistently received higher ratings for instruction on the FFT than those in the bottom quartiles. The difference between being assigned a top-quartile versus a bottom-quartile teacher was associated with a seven-percentile gain in reading and a six-percentile gain in math. Similarly, Newmann et al. (1996) found that authentic pedagogical practices were associated with improved academic performance for students at all grade levels in both mathematics and social studies. A multilevel replication of their study showed that the standards of intellectual quality in authentic pedagogy were associated with improved student performance, above and beyond students' individual prior achievement and social backgrounds (Ladwig, 2007). Schacter and Thum (2004) found that teachers who implemented distinguished or proficient levels of quality according to their model produced students who made considerable achievement gains on the 2001 Stanford Achievement Test. In their study, instructional quality and classroom composition accounted for approximately 84% of the variation in student achievement gains in reading, mathematics, and language. All else being equal, one standard deviation increase in instructional quality translated to increased classroom achievement gains of 10.9 scale points in language, 11.4 scale points in mathematics, and

7.6 scale points in reading, which approached mitigating the effects of students' home environment, prior knowledge, or parental income.

Aguirre-Munoz et al.'s (2006) study of English Learners' access to academic language instruction found that the students in classes with teachers who had knowledge of and taught functional grammar concepts had higher performance than the students in the classrooms whose teachers demonstrated limited to no knowledge of academic language instruction. Similarly, Matsumura et al. (2008) found that both observations and the quality of the tasks teachers gave to students significantly predicted changes in student learning. An important limitation of their study, however, was that the researchers did not use multilevel modeling to account for the clustering of students within classrooms and schools. They used linear regression to estimate their effects, which may have made the effects appear stronger than they actually were because any number of other factors may have come into play to create the results that were obtained. The researchers encouraged further research in larger samples in order to control for the nesting of students within classrooms so that other variables can be more carefully accounted for.

Link Professional Development to Specific Teacher Needs

Traditional professional development activities have been shown to have little effect on improving teacher practice because they are often planned to address general district goals as opposed to the specific learning needs of teachers as they relate to those goals (Newmann, 2000; Steiner, 2004). Kane et al. (2011) found that the detailed descriptions of the range of practice associated with each measure on the FFT rubrics allowed teachers and administrators to thoughtfully consider ways to direct resources

toward improving specific practices through focused professional development. Schacter and Thum (2004) suggested that districts invest in trained professional developers who are able to provide accompanying professional development and modeling of effective teaching practices so that all teachers have equal opportunities to improve their practice.

In response to the observational findings that no teachers met standards of quality for academic language instruction, Aguirre-Munoz et al. (2006) designed a set of training modules and implemented them over a four-day period with a subset of teachers focusing on specific knowledge shown to impact learning for students who are also learning English. They found that the percentages of teachers implementing low and high academic language instructional practices were consistent with whether they were trained or not. The majority of untrained teachers provided no instruction in academic language, and none of them provided moderate or strong instruction. In contrast, nearly 70% of trained teachers provided moderate to strong levels of instruction in academic language after participating in the training. As a result of stronger instruction, the students with access to a teacher with more expertise scored higher on a standardized writing measure.

This underscores the importance of professional development that is differentiated and directly addresses the specific areas of a teacher's practice that need to be strengthened. The data generated from narrative descriptions of teaching scenarios can serve as powerful models for professional development sessions. Elmore (2002) instructed that, "a school system's capacity to make productive use of professional development is directly related to its willingness to make binding and public judgments about quality and expertise" (p. 28). These studies provide empirical evidence that it is possible to assess instruction so that actions can be taken to improve it.

Additional Findings

The research presented in this review demonstrates that classroom observations, when conducted by trained professionals, can successfully identify the teaching practices most likely to improve student achievement and can be integrated into teacher evaluation systems that more effectively assess the instructional contributions effective teachers make toward children's learning. In a follow-up study, Matsumura et al. (2008) confirmed that as few as two observations and the analysis of only four instructional tasks would yield reliable estimates of instructional quality. While this finding supports the feasibility of their approach for measuring instruction at-scale insofar as it imposes a minimum of additional work on the teacher, it also highlights an important limitation. The observation protocol required teachers to adapt their lessons to ensure discussion would be taking place during the observation, so the lesson may not have been a true representation of the teachers' everyday practices. Knowing in advance that an observer was coming in and knowing in advance that discussion would be emphasized may have altered the flow of instruction away from its normal course (Marshall, 2012b). In addition, their protocol for assessing task quality was dependent on teachers submitting both detailed descriptions of the task and a summary and reflection on the task's purpose and grading criterion. Because engaging in these behaviors did not necessarily occur during the normal course of teaching for all educators, the process itself may have skewed the data in a more positive direction than might be the case had the evaluation not been taking place. On the flip side, however, using an evaluation method that prioritizes the processes that have value for students and are consistent with an understood vision of quality instruction may serve to instill in teachers the recognition of the characteristics of

instruction that result in the highest levels of learning. In this instance, “teaching to the test” would be a positive outcome in that teachers would know the criteria upon which they were being judged in advance and potentially strive to achieve high marks for their instructional performance.

The studies also show that all students benefit equally from high-quality instruction. Like Aguirre-Munoz et al.’s (2006) finding that the opportunity to learn functional grammar equally benefited both English language learners and non-English language learners, regardless of the measure of quality, it was shown across studies that the positive efforts of effective teachers had a positive effect on all students. Newton et al. (2010) provided valid evidence that it is possible for schools to provide authentic instruction in an equitable manner and it is possible to construct measures that accurately assess the quality of instruction so that classrooms can be identified where teachers are in need of assistance. Since all students benefit from high-quality instruction, it is important to ensure that all students have access to it.

Improving Instructional Quality

In order to understand whether all children have equitable opportunities to develop their potential, the profession has been challenged to manage the tension between defining instructional quality and assessing the ability of teachers to actually deliver it. As a system, we are still learning about how best to manage accountability so that it has positive effects on school cultures. Hargreaves and Shirley (2009) summarized the history of educational reform in the United States, beginning in 1950. They described three waves of change that were characterized first by significant increases in the allocation of resources, then by strict accountability measures, and finally by a test-driven

devotion to developing basic skills in math and reading at the expense of high-quality, rich, engaging instruction across all content areas (Kohn, 1999).

The first wave of educational reform was fuelled by the belief that, once the government apportioned resources to education, professionals could be trusted and left alone to get on with their job, without interruption or intervention (Hargreaves et al., 2009). Although allocation of resources was an important first step in addressing critical problems, it was not sufficient for ensuring effective implementation of learning goals. For example, having enough teachers is critically important, but understanding how the work they do impacts student learning requires a higher level of attention and more focused professional accountability (Darling-Hammond, 2010; Ladwig, 2009; Olson, 2003). A lack of vision and cohesion in regard to the instructional quality that was expected of teachers resulted in huge variations in both content and pedagogy and led to “inconsistent performance, unpredictable leadership, and educational improvements informed by intuition and ideology rather than through evidence-informed initiatives” (Hargreaves et al., 2009, p. 9).

In response to the lack of a commonly held vision, the second wave of school reform produced an over-compensated preoccupation with control and accountability. In this environment, innovation gave way to standardization and the belief that reforms that were initiated at the top would be implemented with fidelity. In many instances, teachers were expected to follow very rigid guidelines and ignore any knowledge or expertise they had developed previously. The impact was a severe decline in quality of instruction, teacher motivation, leadership capacity, and student learning (Hargreaves, 2012).

In the third wave of education reform, the environment proved to be even more autocratic, more centered on accountability, and increasingly more intrusive in terms of focusing instruction almost exclusively on developing basic skills in English and math in an effort to increase student achievement (Hargreaves, 2012). In this wave, it was believed that all achievement gaps could be detected from data, and districts scrambled to figure out how best to use the data to influence higher test scores. As schools and school systems emphasized bureaucratic accountability in the form of management of test scores at the expense of professional accountability in the form of instructional efficacy, teachers' ownership of the instructional process decreased significantly (Darling-Hammond, 2010; Guskey & Passaro, 1993; Hargreaves, 2012).

Changing practice on a large scale means that school leaders provide necessary foundational conditions: a clear, inspiring, and co-created vision of excellence; a shared understanding of rigorous and relevant standards and assessments; clear, agreed-upon expectations and differentiated support for teachers and students; and the core resources needed to build the knowledge, skills, strategies, and commitment of teachers over time (Barber & Mourshed, 2007; Schein, 2010). The profession has already revised standards to reflect 21st century learning. Many schools have structures in place that allow teachers to reflect on and collaborate to improve instruction, but school-level visions of what instruction, collaboration, and reflective practice look and feel like need to be more carefully understood and articulated. In order to develop individual and group capacity to improve instruction, a theory of action is needed that identifies specific standards against which to observe and measure a teacher's effort, performance, and accomplishments

without stifling teachers' abilities to provide creative and innovative instruction (Smylie, 1995).

Kaplan (1964) defined a theory as a set of logically interconnected concepts that are operationally defined and testable in the universe they purport to represent, but Argyris and Schön (1996) augmented that definition by adding that it requires people to take action, which generates the creation of new ways of behaving that produce increasingly more effective results. The chart in Figure 1 redistributes the research-based components of effective teaching that have been described in this review into three distinct areas of teacher expertise. The highly inter-related nature of the model is demonstrated in the subsequent rows, where the overlap is evident. The ten qualities in the last section, reflective practice, significantly influence, and are influenced by, all three domains.

Domains of Expertise for Quality Instruction		
Content Area Expertise	Pedagogical Expertise	Classroom Culture
Knowing the standards Knowing the curriculum Setting instructional outcomes Explaining concepts Make decisions about pacing and concept emphasis Integrating textbooks, teacher's guides, materials, and manipulatives Expecting students to justify explanations Analyzing data Determining grading criteria	Knowing strategies Using questioning and discussion techniques Demonstrating flexibility and responsiveness Eliciting student participation Pacing lessons to maintain engagement Expecting students to raise key questions	Knowing students Creating norms of collaboration Managing classroom procedures Establishing efficient classroom routines Communicating with students Building cohesiveness among students Maximizing time on task Maintaining accurate records Communicating with families Managing student behavior Ensuring student safety
Interaction between Content Knowledge and Pedagogy Planning engaging and challenging lessons and integrating a variety of resources Designing coherent instructional tasks Implementing clear and organized direct instruction Articulating clear standards for student work Designing student assessments that are integrated with instruction Frequently calling for extended, substantive oral and written responses Facilitating opportunities for students to construct their own knowledge Expecting mastery of core concepts and checking for understanding Expecting students to solve rich problems Expecting students to synthesize, apply and interpret concepts Intentionally equipping students with metacognitive routines for understanding, monitoring, and self-correcting comprehension errors Interpreting and representing student thought processes Summarizing learning and providing lesson closure Bridging learning to more complex understandings		
	Interaction between Content Knowledge and Culture Building on students' prior knowledge Creating appropriate academic goals Instilling appreciation for learning Monitoring progress toward goals Representing concepts in a variety of ways Facilitating connections between concepts Creating many opportunities for students to use higher-order thinking Expecting students to substantiate claims and arguments Scaffolding academic language Giving students many opportunities to explain, clarify, and challenge ideas Using formative assessment during instruction Re-explaining concepts with analogies, metaphors, examples and demonstrations Understanding what makes learning the concept easy or difficult	
	Interaction between Pedagogy and Culture Establishing a culture for learning Co-creating, communicating, and enforcing classroom rules Organizing physical space Understanding, leveraging, and fostering high-quality student motivation Encouraging students to take risks Grouping students to foster a productive use of instructional time Creating an environment of respect and rapport Using questioning/recitation strategies that maintain active participation (group alerting)	
Reflective Practice (Praxis) Engaging students actively in intellectual inquiry Scaffolding Instruction Differentiating instruction Integrating and using technology Reflecting on learning and evaluating effectiveness Reflecting on teaching and evaluating effectiveness Taking appropriate responsibility for student success or failure Adapting practice to address student needs Showing professionalism Participating in a professional learning community		

Figure 1. An interactive view of effective teaching

Reflective practice or “praxis” is the idea that teachers will continually engage in a process of self-analysis of one’s knowledge of content as it relates to children’s thinking about the concepts they teach (Silver, Ghouseini, Gosen, Charalambous, & Strawhun, 2005). It represents a teacher’s capacity to analyze their practice, identify and wrestle with learning dilemmas, and engage in active reflection that results in actions that are different than what would have occurred otherwise (Boyd & Fales, 1983; Schön, 1983). It involves engaging in a collaborative cycle of theorizing, applying, evaluating, reflecting, and then back to theorizing to continually evolve in one’s ability to refine teaching (Argyris, 1993; Argyris & Schön, 1996; Schein, 2010).

A Theoretical Framework for Instructional Quality and Measurement

The observation instrument used in this current study was developed from the theoretical framework presented in Figure 2 and was designed to achieve an integrated measure of instructional quality to help guide and facilitate the instructional improvement process. The theory first defines the attributes of effective teaching as described in the three circles making up the Venn diagram (see Figure 2). Next, it integrates known student-related factors that support student engagement including rigor, relevance, and relationships (Daggett, 2002). The novelty of this model is the specification of three critical variables to observe in students as instruction is taking place: participation, critical thinking, and academic language. Finally, it aligns with Elmore’s (2009) recommendations for improving instructional quality, which are added to the outside edges of the triangle. The prominent position of the teacher as the major force driving the instructional core highlights the fact that it is the teacher’s knowledge, beliefs, and dispositions that dictate the quality of tasks they create that either engage or disengage

students in active learning (Cohen & Ball, 1999; Ferguson, 2003; Borko, 2004). The instructional task takes center stage in the model and embodies the teacher's ability to enact the vision of instructional quality (Bowen, 2003).

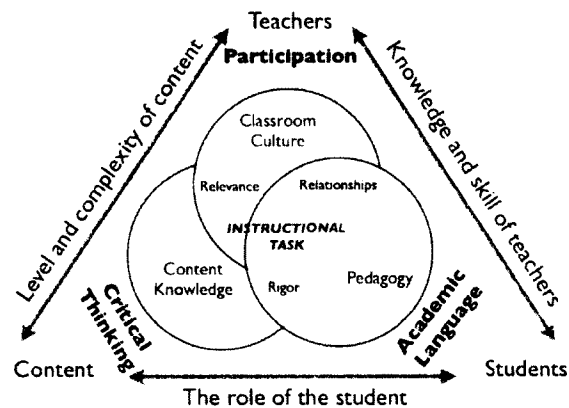


Figure 2. A theoretical framework for instructional quality and measurement

Content Knowledge, Pedagogy, and Classroom Culture

The theory is based on the idea that high-quality instruction is a complex and dynamic interchange that it is dependent upon skill sets developed along three overlapping and interconnected dimensions: (a) deep knowledge of content and curriculum (Grossman, Stodolsky, & Knapp, 2004; Hill et al., 2005; Krauss, et al., 2008; Porter, 2002), (b) strong pedagogical skills that promote student engagement (Bowen, 2002; Corno & Mandinach, 1983; Delpit, 1988; Kong & Hoare, 2011; Tyler & Boelter, 2008), and (c) the ability to create a classroom culture centered on substantive, interactive learning (Dooner et al., 2010; Lounsbury & Clark, 1990; Sternberg, 2003; Tishman, Perkins, & Jay, 1995)

Rigor, Relevance, and Relationships

Regarding the student perspective, recent research on the racial achievement gap, dropout prevention, and school reform emphasizes the importance of rigor, relevance,

and relationships (Daggett, 2004; Picower, 2009; Resnick & Glennan, 2002; Rumberger & Lim, 2008; Sofo, 2008). Rigorous instruction is created when a teacher's content knowledge interacts with effective pedagogical skills and results in tasks that challenge students to think creatively and solve complex problems (Bowen, 2003; Bransford et al., 1999; Dooner et al., 2010; Resnick & Hall, 2001; Smith & Geller, 2004). Relevance is present when the teacher uses his or her knowledge of students along with his or her knowledge of content and curriculum to craft tasks students understand, enjoy, are interested in, and take seriously (Bowen, 2003; Costa, Kallick, & Association for Supervision and Curriculum, 2008). Relationships are strengthened when a teacher's pedagogical skills interact with a positive classroom culture which values, respects, and appreciates students' contributions to the teaching and learning process (Cohen et al., 2003; Davis, 2006; Haberman, 1992; Nystrand, 1997; Olson, 2003; Pang, Stein, Gomez, Matas, & Shimogori, 2011; Resnick & Hall, 2001; Wentzel, 1997). When rigor, relevance, and relationships coalesce, the teacher achieves a state for students that Shernoff and Csikszentmihalyi (2009) defined as flow. Flow is a state of concentration and complete absorption with the activity where students are intrinsically motivated to engage in the task. Flow happens when a balance is struck between the challenge of the task and the skill of the performer. If the task is too easy or too difficult, flow cannot occur and frustration or apathy is the result. This leads to reduced student engagement and, over time, the potential for students leaving school with the belief that formal education is not an endeavor that is worth their effort or attention (Donner et al., 2010; Sofo, 2008).

Participation, Critical Thinking, and Academic Language

Porter (1991) pointed out that, “conceptually, it may be impossible to define student perception (or teacher perception) of content taught, separately from student achievement, but the perspective of an external observer may be the best measure available” (p. 18). The integrated theoretical model of instructional quality presented here suggests that the components of professional practice, which relate to high-quality instruction in traditional teacher evaluation rubrics (Danielson, 1996, 2007; Schacter & Thum, 2004), may be as effectively measured by watching students, as it is by watching teachers. A teacher’s ability to facilitate multi-directional interactions with the class that result in authentic cognitive engagement and increased subject matter competence (Cohen, Raudenbush, & Ball, 2003; Goe & Stickler, 2008; Matsumura et al., 2006; Tyler & Boelter, 2008) can be achieved in a variety of ways. Because very successful teachers often display quite different characteristics of effectiveness, it makes sense to focus on observable student actions that result from teacher decision-making rather than focusing solely on the actions of the teacher. The observable student actions suggested by this model include participation, critical thinking, and students’ use of academic language.

Student participation during a lesson serves as a proxy for evaluating the quality of the classroom culture. When students participate actively in the learning process, it reflects a teacher’s ability to motivate and communicate with students, create an environment where students feel safe and respected, and organize the classroom for positive student interaction and learning (Boettcher, 2007; Dooner et al., 2010; Flynt & Brozo, 2009; Graff, 2009; Hiebert & Grouws, 2007; Matsumura et al., 2008; Porter, 1995; Schacter & Thum, 2004).

Critical thinking serves as a proxy for evaluating a teacher's content-area expertise. A teacher's deep and broad content-area expertise enables the teacher to design developmentally appropriate learning tasks, set challenging and differentiated instructional outcomes, effectively use and integrate a variety of resources, and represent concepts in a variety of ways (Bowen, 2003; Dooner et al., 2010; Krauss et al., 2008; Nasstrom, 2009). Deeper understanding of critical thinking began with the introduction of Bloom's Taxonomy in 1956, which categorized thinking from simple to complex, and from concrete to abstract. Mastery of each simpler level was a prerequisite for mastery of more complex levels. Recognizing the importance of both the distinction and the interactions that exist between knowledge and cognition, Anderson, et al. (2001) restructured the original version of Bloom's Taxonomy to create a two-dimensional framework. The revised framework has since been integrated with depth of knowledge (Webb, 2005) levels which include recalling information, applying skills and knowledge, thinking strategically, and extending student thinking. Webb (2005) developed the depth of knowledge levels to facilitate the analysis of whether or not standardized assessments were consistent with the content standards they claimed to measure (Webb, 2007). The detailed descriptors at each level include both the content and the depth to which we expect students to demonstrate understanding of that content (Hess, et al., 2009).

Teachers with expert content knowledge are able to use research-based instructional strategies to develop student's academic competence and critical thinking about subject matter (Perkins, Jay, & Tishman, 1993; Kannapel & Clements, 2005; Raudenbush & Ball, 2003). Teachers with greater expertise in their content area have been shown to cover more content with greater effectiveness (Boacardin et al., 2005).

The extent to which teachers engage students in critical thinking is dependent on the tasks they design and the scaffolds they put in place that allow students at varying levels of ability to use higher-order thinking skills to make meaning of content (Anderson et al., 2001; Bowen, 2003; Hatano & Osuro, 2003; Pang et al., 2011; Sawyer, 2006; Schulman, 1986).

Students' use of academic language serves as a proxy for evaluating a teacher's pedagogical skills. The quantity and quality of language as an observable factor emphasizes the importance of children actively constructing meaning of content through the dialogic use of language (Bakhtin, Holquist, & Emerson, 1981; Cazden, 2001; Halliday, 1993; Soter, et al., 2008). Rather than the traditional three-part exchange, in which the teacher asks a question, a student is selected to answer, and the teacher evaluates the student's response, known as initiation-response-evaluation or IRE (Mehan, 1979), effective teachers facilitate multi-directional conversations where students articulate, examine, elaborate, and revise their developing ideas about content (King, 1994; Nystrand, 1997; Resnick & Hall, 2001). Effective pedagogy is distinguished from ineffective pedagogy by the teacher's ability to adapt the initial question so that all students as opposed to one or just a few think about the answer and his or her ability to respond on the third turn in a way that engages the class in critical examination of the concept (Lee, 2007). These actions cause students to engage in multi-directional discussions that integrate academic vocabulary. These exchanges have been shown to more effectively develop problem-solving, reasoning, and communication skills, and cause students to experience higher levels of individual academic achievement (Halliday, 1993; Snow, Lawrence, & White, 2009; Westgate & Hughes, 1997).

Each of these variables by itself provides only a snapshot of the quality of instruction a teacher provides, but observed together throughout an entire lesson, they allow the evaluator to track teacher actions that maximize or undermine student engagement in learning. When the three variables are combined, their collective value represents the teacher's ability to enact the expectations of high-quality instruction (Aguirre-Munoz et al., 2006; Goe & Stickler, 2008; Hiebert & Grouws, 2007; Matsumura et al., 2008; Milanowski, 2011; Newmann, et al., 1996; Resnick & Hall, 2001).

Instructional Improvement

Barber and Mourshed (2007) studied 25 of the world's school systems, including 10 of the top performers, to find out why some school systems consistently perform better and improve faster than others. They concluded that the quality of an education system cannot exceed the quality of its teachers. The only way to improve student outcomes is to improve teacher instruction (Barber & Mourshed, 2007). Elmore (2009) proposed that the improvement of instruction is achieved in one of three ways: (a) by increasing the knowledge and skill of teachers, (b) by increasing the level or complexity of the content, or (c) by significantly altering the role of the student in learning. These improvement avenues are represented along the outside edge of the triangle, suggesting that the data collected from the observation will be used to determine next steps for the teacher's own learning. For example, if participation was low, then professional development would focus on re-evaluating the role of the student in instructional tasks (Bowen, 2003). If critical thinking needed strengthening, then professional learning would highlight increasing the level and complexity of the content (Anderson et al., 2001; Reginier, 1994). If academic language was not integrated into the development of

student understanding, then professional development would aim to strengthen the teacher's pedagogical skills so that academic language can be more effectively scaffolded. Elmore (2002) has instructed that a central part of the discipline of improvement is the belief that if the teaching is good and powerful, and if the conditions of work enable and support strong practice, then we should be able to see immediate evidence that students are learning. If we cannot, then we should ask if the teaching is really as good as we thought it was.

Teachers cannot be expected to accept responsibility for their instructional outcomes unless they are granted the autonomy to exercise individual judgment in meeting them (Olson, 2003). When student outcomes are the focus of an observation, the creativity and art of teaching is effectively captured, but not constricted, by the data collection and measurement process. The data produced by the proposed instrument allows the teacher to see where student engagement was high and in what parts of the lesson students stopped paying attention or being engaged in substantive thinking and learning. The evaluation is not a prescription for how the teacher should teach, but rather provides data about what students did as a result of how the teacher taught. With this focus, teacher autonomy is maximized. As such, teachers are equipped to engage in reflective thinking about their instruction because the adult learning cycle is supported by data that is both valid and reliable, which helps them to be analytical about how they can improve (Goe, Bell, Little, & National Comprehensive Center for Teacher Quality, 2008; Parsons & Brown, 2002; Patton, 2008). Noticing aspects of the instruction that fell below the expected threshold for effectiveness informs teachers and school leaders about where to focus professional development and improvements efforts (Fink & Resnick, 2001).

Accelerating and sustaining this cycle of continuous improvement was the primary goal for this study.

CHAPTER THREE: METHODOLOGY

This parallel convergent mixed-methods designed study utilized quantitative measures in combination with qualitative analysis to evaluate instructional quality for the purpose of measuring and improving teacher practice. This chapter presents the research questions and the procedures for collecting, measuring, and analyzing data.

Research Questions

The following research questions were designed to assess the impact of instruction on student achievement measures and also to determine what attributes of instruction yield the highest and lowest levels of instructional quality. Two main research questions guided this study. The first, which is divided into three sub-questions, guided the quantitative analysis, while the second question was answered using qualitative research methods.

1. What is the impact of instructional quality on student achievement?
 - a) How much do classrooms vary in the average level and shape of student growth trajectories in mathematics and English/Language Arts over the course of a school year?
 - b) How much of the difference in student growth trajectories is accounted for by student factors beyond the teacher's control?

- c) Is the variance between classrooms and teachers in levels of participation, critical thinking, and students' use of academic language associated with the average level and shape of student growth trajectories?
2. What instructional practices yield the highest and lowest levels of instructional quality?

Research Context

This study took place in two California middle schools. Middle school represents a critical time in a student's academic career because it lays the foundation for success in high school and beyond. One of the schools has struggled to exit program improvement, which means it has not made adequate yearly progress for five consecutive years, based on the requirements of No Child Left Behind (2001). The other school has shown continuous improvement over the last five years. Including both schools increased the likelihood that the data would reflect a broad range of teacher practice and student performance. This inclusion is significant because the success of the children who attend lower performing schools is more likely to be influenced by the quality of the classroom instruction they receive than students attending higher performing schools who have greater access to other resources. As such, including a broad range of quality helps illuminate the impact of those differences (Berry, 2010).

Including two different schools also allowed for deeper understanding of the differences between instructional supervision which is conducted by a person in a position of power for the purpose of evaluation, and instructional observation outside of the evaluation process, which occurs for the purpose of instructional reflection and improvement (Mertens, 1999; Snow-Gerono, 2008; Western, 2008). Each of the

classroom observations were conducted by the researcher. In one of the schools, she was in the role of vice principal and in charge of teacher evaluation, course assignments, professional development, and student discipline. In the other school, she was a researcher only, with no other responsibilities in the school. Analyzing the variance in instructional quality across the two schools has the potential to inform the dialogue concerning who should conduct classroom observations so it was important that the range of data included classroom observations taken in both contexts.

Research Design

Mixed methods were used in all three phases of the study including the collection, analysis, and interpretation of data. Simultaneous analysis and the integration of quantitative and qualitative datasets provide a more holistic description of complex phenomena (Caracelli & Greene, 1997; Creswell & Plano Clark, 2011). The observation instrument used in this study includes numerical measures of instructional quality alongside narrative descriptions of instructional tasks. The analysis required both multilevel modeling techniques (Heck et al., 2010) and the extrapolation of patterns from comparative analysis (Patton, 2008; Stake, 2004) to discern the relationship between student achievement and classroom instructional quality as well as to describe the attributes of instruction that yielded the highest and lowest levels of student engagement.

Sample

A total of 2594 students participated in this mixed-methods study, including 794 sixth graders, 739 seventh graders, and 1061 eighth graders. The students were clustered in 54 math classrooms and 59 English/language arts classrooms at two middle schools in an urban district in California. The diverse student body included 35% Hispanic, 22%

Filipino, 19% White, 13% African American, 10% Asian (not Filipino), and 1% other, including, Pacific Islander, Guamanian, and Native American. Sixty percent of students qualified for free or reduced-priced meals, 21% were English Learners, and 15% comprised students with disabilities. In the 2011–2012 school year, the first school had a total of 465 suspensions (second highest in the district for middle schools) and the school's Academic Performance Index (API) score was 738. API is a state-wide ranking based on the academic performance and progress of individual schools that ranges from a low of 200 to a high of 1000. Seventy-one percent of students at the first school qualified for free/reduced priced meals. The second school had a total of 77 suspensions and an API ranking of 887. Forty-eight percent of students at this middle school qualified for free/reduced price meals.

The Dependent Variable

The dependent variable in this study is each student's individual achievement trajectory, which was calculated from their 2011 California Standards Test (CST) scaled score and their 2012 CST scaled score. The total data set had 1659 students in math and 1887 students in English. The listwise (i.e., students with no missing data) was 1459 (88%) in math and 1646 in English (87%). This would represent a considerable amount of missing data that would likely bias the results unless it could be demonstrated that the data are missing completely at random (MCAR). It is more common that data will be either missing at random (MAR) or missing not at random (MNAR). Handling missing values appropriately depends on having some knowledge of the data set and why particular values may be missing (Heck, Thomas, & Tabata, 2010).

There are currently two acceptable ways for dealing with missing data recognized in the literature—multiple imputation of plausible values and full maximum likelihood estimation in the presence of the missing data (Peugh & Enders, 2004; Rubin, 1987). Where the probability of missing data on an outcome such as achievement may be associated with a covariate, but not with subjects' standing on the covariate (i.e., where students who have missing data on attendance do not have a greater probability to have low scores in math), we can assume the data are MAR. Where the probability of the data being missing is related to subjects' standing on the outcome even for those with the same attendance level (i.e., there are more missing math data for those with low attendance than average or high attendance), then the data are likely MNAR. If the data are missing at random, then either multiple imputation or maximum likelihood estimation in the presence of missing data can be used to estimate the models.

Including individuals with partial data in the analyses is important for the assumption of MAR (Peugh & Enders, 2004). Fortunately, where data are vertically arranged in SPSS (Statistical Package for the Social Sciences), as in the present analyses, individuals with partial data are included. As might be expected, some students had partial test score data for either 2011 or 2012, and a small percentage had missing data for both years. In this case, it was determined that of the 1659 students in math, 54 individuals had partial data and were included in the final analyses (i.e., with only 4.6% being missing on both years). Of the 1887 students in English, 67 had partial data and were included in the final analyses (with 6.0% missing data on both years).

After this examination, then, 1583 participated in math and 1773 students participated in English. This represented 95.4% of the total sample in math and 94% in

English. The math students were from 54 classrooms nested within 15 teachers. The English students were from 59 classrooms nested within 18 teachers. In order to determine whether there were differences in the initial sample and the final sample after applying the complete-test criterion, several follow-up analyses were conducted.

Preliminary Analysis of Missing Data

It is important to determine whether the missing data will likely bias the estimation of the model parameters. For example, when only the individuals with complete data are considered, it is possible that achievement means for the classes would be over-estimated if the cause for missing data is high absenteeism. Results using maximum likelihood estimation with partial data included and multiple imputation can be compared for consistency. Rubin's approach to multiple imputation involves estimating a number of "imputed" data sets. Then the simulated complete datasets are analyzed by standard methods, and the results are combined to produce estimates and confidence intervals that incorporate missing-data uncertainty. The approach borrows information from the covariates used in the analysis, as well as from complete cases on the variable where imputed values are desired. Five imputed data sets were developed to test the missing data in this where plausible values for students missing 2011 or 2012 math and English scores were randomly imputed to generate complete data sets with student demographics included in the models (Rubin, 1987).

In Tables 1 and 2, the pooled results across all five analyses are presented. In each of the five separate data sets generated in math and English, the effect of whether a student was "missing" from the initial data set or had only partial outcome data had no significant effect on the estimated outcomes after controlling for the students'

background data. For comparative purposes, the listwise math intercept, controlling for student background variables, was 372.58 (not tabled). For English, again controlling for background variables, the listwise estimate was 373.35 (not tabled). Because the intercepts with multiple imputation were quite similar to the listwise data, where individuals with any missing values on the outcomes were eliminated, the data are likely MAR and, therefore, not connected to the presence of background variables such as individual absenteeism or behavioral patterns. This provides one baseline against which to compare results estimated with individuals with partial data included in the model.

Table 1

Estimates examining the effect of missing data on 2012 math outcomes

Parameter	Estimate	Std. Error	T	Sig.	95% lower	95% upper
Intercept	372.56	9.26	54.87	0.00	358.66	386.46
Missing flag	-0.77	5.36	-0.14	0.89	-12.14	10.59
EL	-28.02	3.06	-9.15	0.00	-34.08	-21.97
AA	-15.41	3.91	-3.94	0.00	-23.13	-7.69
HISP	-13.61	3.21	-4.24	0.00	-19.96	-7.25
Filipino	5.46	3.77	1.45	0.15	-1.95	12.87
OtherAsian	11.30	4.21	2.68	0.01	2.96	19.65
Gate	34.78	2.68	12.97	0.00	29.49	40.06
ZAttendance	-12.00	1.59	-7.55	0.00	-15.42	-8.57
ZBehavior	-6.52	1.87	-3.49	0.01	-10.79	-2.26
ZSESmean	-21.92	3.50	-6.27	0.00	-28.77	-15.07

Yearly gain	-16.91	2.01	-8.43	0.00	-20.86	-12.97
-------------	--------	------	-------	------	--------	--------

Table 2

Estimates examining the effect of missing data on 2012 English outcomes

Parameter	Estimate	Std. Error	T	Sig.	95% lower	95% upper
Intercept	373.29	3.66	102.03	0.00	365.94	380.65
Missing flag	1.11	2.91	0.38	0.70	-4.75	6.97
EL	-39.36	2.05	-19.25	0.00	-43.37	-35.35
AA	-25.32	2.88	-8.78	0.00	-31.10	-19.55
HISP	-18.22	2.26	-8.05	0.00	-22.71	-13.72
Filipino	-0.62	2.77	-0.22	0.82	-6.09	4.85
OtherAsian	-1.79	2.87	-0.62	0.53	-7.43	3.84
Gate	22.35	1.92	11.62	0.00	18.56	26.14
ZAttendance	-7.34	1.05	-6.96	0.00	-9.56	-5.11
ZBehavior	-7.59	0.96	-7.88	0.00	-9.57	-5.61
ZSESmean	-11.81	1.65	-7.18	0.00	-15.04	-8.59
Yearly gain	0.54	1.42	0.38	0.71	-2.26	3.33

In contrast, maximum likelihood estimates can also be calculated directly from the incomplete data where the data are vertically arranged and at least one value is present. These procedures are sometimes more efficient than multiple imputation because they involve no simulation (Puegh & Enders, 2004). The results of the imputed data sets were compared against the maximum likelihood estimates for the data sets. In this case,

the results in Tables 3 and 4 match very closely the results obtained through using multiple imputation across five simulated complete data sets. For example, both of the sample intercept estimates fall within the 95% confidence intervals of the five imputed data sets, as do the regression coefficients on each predictor. In both tables, the effect of individuals with partial data had no significant influence on the estimates of student achievement in either English or math ($p > .10$). A series of interactions was also investigated (not tabled) for each background variable and the missing/partial data indicator, and each interaction was found to be not significant ($p > .05$). This finding suggests that the effect of the individual background predictors on the outcomes was not different for individuals with complete versus missing data. The findings in Tables 3 and 4 provide evidence that any data missing on outcomes (which is about 5%) can be considered MAR and should not bias the mode estimates (Hox, 2010).

Table 3

Estimates of math fixed effects for missing data

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	362.71	9.02	563.96	40.20	0.00	344.99	380.43
EL	-27.82	3.03	2892.65	-9.19	0.00	-33.75	-21.89
AA	-15.38	3.80	2888.48	-4.05	0.00	-22.84	-7.93
HISP	-13.12	3.03	2890.88	-4.33	0.00	-19.06	-7.18
Filipino	5.23	3.73	2905.53	1.40	0.16	-2.09	12.55
OtherAsian	9.94	3.86	2866.25	2.58	0.01	2.38	17.50

Gate	33.05	2.49	2906.04	13.28	0.00	28.17	37.92
ZAttendance	-11.41	1.31	2869.80	-8.71	0.00	-13.97	-8.84
ZBehavior	-5.51	1.29	2871.28	-4.28	0.00	-8.03	-2.98
Gain	-18.32	1.99	2854.90	-9.21	0.00	-22.22	-14.42
Missing	10.37	7.52	2862.51	1.38	0.17	-4.37	25.11

Note: a. Dependent Variable: mathscore.

Table 4

Estimates of English fixed effects for missing data

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	366.62	5.75	1464.02	63.73	0.00	355.34	377.91
EL	-39.45	2.21	3280.33	-17.84	0.00	-43.79	-35.12
AA	-24.99	2.68	3279.54	-9.32	0.00	-30.25	-19.73
HISP	-18.23	2.15	3280.77	-8.49	0.00	-22.44	-14.02
Filipino	-0.32	2.71	3095.13	-0.12	0.91	-5.63	5.00
Other Asian	-1.25	2.91	3247.82	-0.43	0.67	-6.96	4.46
Gate	21.99	1.86	3163.93	11.82	0.00	18.34	25.64
ZAttendance	-6.94	0.94	3253.59	-7.37	0.00	-8.79	-5.10
ZBehavior	-7.11	0.96	3259.18	-7.41	0.00	-8.99	-5.23
Gain	0.61	1.45	3222.12	0.42	0.67	-2.23	3.45
Missing	6.99	5.20	3235.59	1.34	0.18	-3.21	17.20

Note: a. Dependent Variable: Englishscore.

Data Collection

Data was collected from each classroom in an unannounced full-lesson observation that was representative of the typical instruction provided by that teacher on a daily basis. If the teacher reported on initial arrival in the classroom that the instruction was atypical (for example, if the teacher was ill or the class was going on a fieldtrip) or if a substitute was present, then the observation was cancelled. After the observations were conducted, teachers were provided feedback on the quality of their instruction in terms of the three dimensions: participation, critical thinking, and students' use of academic language.

Instrumentation

The instrument presented in this study was developed collaboratively with teachers while they worked together to identify the critical factors that one observes when powerful instruction is taking place. It is unique because it focuses on observable student behaviors that happen as a result of teacher decision-making. It provides both a quantitative measure of instructional quality and qualitative descriptions of instructional tasks so that a lesson's quality can be both measured and understood. The quantitative component is based on factors directly influenced by the environment teachers create in their classrooms and the instructional decisions they make prior to and during the course of instruction. The quantitative component measures how many students are participating in the lesson, at what level of critical thinking students are engaged throughout the lesson, and the amount of academic language that students use as they interact with the concepts they are learning. Because determining value is always bound by context (Stake, 2004), the qualitative component of the instrument requires the observer to describe the

instructional task. This description includes the questions the teacher asks, descriptions of student work, student responses during instruction, and interactions between students and the teacher and students and other students (Nystrand, 1997).

The observation protocol requires 10 lesson segment observations throughout the course of a 50-minute class period at approximate five-minute intervals. The participation score represents the number of students who were doing what the teacher expected (0=none, 1=a few, 2=about half, 3=most, 4=all). The critical thinking score corresponds to Webb's (2005) *Depth of Knowledge* levels (0=no thinking, 1=recall, 2=skill/concept, 3=strategic, 4=extended). The academic language score measures the quantity and quality of expressive language required by the task (0=no language, 1=one-word answers or phrases, 2=sentences, 3=phrases or sentences with academic language, 4=multi-directional sentences with academic language). The instructional quality measure reported for each teacher reflects the total score for participation (0–40), critical thinking (0–40), and academic language (0–40) elicited from students by the teacher during instruction for a total instructional quality score that ranges anywhere between 0 and 120 (Please see Figure 2).

Performance Assessment for Quality Teaching (PAQT)									
Teacher	Grade	7	Observer	Quality Indicators					
Date	3/18/11	# of Students	20	Formal or Informal?	Formal	Formal	Formal	Formal	Formal
Subject	PreAlg	Period	3	Observed Date	3/18/11	1	2	3	4
<p>Graph of Participation, Critical Thinking, and Academic Language</p> <p>Time (Approximate 5 Minute Intervals)</p>									
#	Time	P	CT	AL	T	Instructional Task			
1	945	4	0	0	4	Prep Talk before students enter the room. Today we are going to review for the test we are taking tomorrow. It's important that we work together so we can all achieve the highest score possible. That means you need to be ready to think and ask questions. Grab your your notebooks, join me at the rug, and write the six key concepts covered in chapter 6 on the next blank page (displayed for copying).			
2	950	4	3	1	8	Now, look back at your notes, thinking about each of the concepts: Ratios, Rates, Proportions, Decimals/Fractions/Percent.			
3	955	2	3	4	9	Discount/Mark-Up, and Point of Change. Rank each of the six concepts from easiest to hardest (1 - 6).			
4	1002	4	2	4	10	Teacher surveying class to get more detailed information. It looks like everyone is still confused about rates and unit rates. Made sure everyone had a partner. Explain to your partner: What do you understand right now about the difference between a rate and a unit rate? Whole class back together. Who can explain to the class how these two concepts relate to each other? S4: A rate is like 50 miles in five hours and a unit rate is like five miles in one hour. I remember it's always one in the denominator. What does it look like visually? Everyone: A fraction. Right. T writing down what student said on P-Board. So if you travelled 50 miles in 5 hours, how many miles did you travel in one hour. Work with your partner and use the numbers to explain the difference.			
5	1007	4	2	3	9	Checking for understanding: Mr. Torres, explain the difference between a rate and a unit rate. "We did miles per gallon, a rate would be 200 miles on 10 gallons and a unit rate would be 20 miles on one gallon." Good! Now let's look at this one: Is it a rate or a unit rate? 100 heartbeats/2 minutes, tell your partner. Everyone: Rate, or unit rate? Rate. How do you know? The denominator is more than one.			
6	1010	4	2	3	9	Pointing to 100/2. Let's turn this into a unit rate. Everyone do your own. Students working. A few knew immediately. Others doing division or reducing the fraction. Glance-You have to divide 100 by 2, right? Teacher nods. What's the final answer? 50/1 50 what over 1 what? 50 beats per one minute!			
7	1015	4	4	4	12	Now I want you to think by yourself of a rate that makes sense to you, then solve it as a unit rate. Some students working. Some thinking, some are stuck. If you are stuck, come over here and we'll work it out together. 7 students came over. Does someone have an idea for a rate? Hot cheetos. Group worked quietly w/ teacher on hot cheeto example while the rest of the students worked on their own. As students finished, they were challenged to explain their idea at their table groups and compare models to make sure they were correct.			
8	1022	3	3	4	10	Now everyone turn to your partner and explain what you did. PT and pointing to their work. Who wants to share? Ms. Dean: I had 120 plates and 80 people. Come show us how you turned that into a unit rate. Student brought notebook up to the doc cam. I divided 80/120 and I got 1/3 plates. Hmmm? Does that make sense? Can you have 1/3 plates? Talk to your partner about this and see what you think. Class confused. Let's do it together.			
9	1028	3	4	4	11	What is the rate? 120 plates/80 people. What is the unit rate? 1/3/1. PT about what that means. S1: I think you have one and a half plates for every person. S2-How does a person have one and a half plates? S3: I think there are three plates for two people (it's the same, right?) So they each have their own and then they have to share the other one. S1-Yeah like they can use that one for cake. Ms. Dean, can you explain what it means to you? Okay, there is one plate for every person. That's 80, and then everyone has to share their second plate with someone. That's the other 40. T-So what was the rate? 120/80. And the unit rate? 1/2/1. See if your partner understands now. PT			
10	1035	4	3	4	11	Now I want everyone to summarize in your own words the difference between a rate and a unit rate. Write it down in your notebook and give an example. Just before bell-partners shared their summaries and one student shared out: "A rate is a fraction, that's denominator is more than one and a unit rate always has a 1." What do we call a fraction that compares two quantities? A ratio (almost everyone).			
TOTALS 36 26 31 93						Quality Indicators		For Total Score:	
The numbers used in this instrument represent the quality of instruction and not the quality of the person. The numbers are meant to be used for reflection on practice so that the quality of instruction can be continually improved.						For each Domain: 90 - 100 Distinguished 80 - 90 Adequate 70 - 80 Requires Development 60 - 70 Inadequate		Summary: 90 - 100 Distinguished 80 - 90 Adequate 70 - 80 Requires Development 60 - 70 Inadequate	

Figure 3: PAQT observation form

The sample observation instrument or Performance Assessment for Quality Teaching (PAQT), presented in Figure 3, summarizes a lesson taught in a seventh grade pre-algebra class. The teacher was presenting a review lesson the day before a test. The instrument allows the observer to assess instructional quality for each five-minute segment as the lesson progresses and also in each dimension overall. The description of the instructional task attempts to capture teacher actions that produce the levels of participation, critical thinking, and academic language that are recorded. Segment seven in this lesson provides an example of the highest level of instructional quality, where all students were participating, thinking critically, and using academic language in their learning.

Reliability and Validity

Because the PAQT is a new construct, it will require more global use in order to assess validity and reliability, but some informal tests have yielded promising results. A reliable observation instrument yields stable results and is consistent in its measurement (QMSS, 2003). It also produces information that provides an accurate description of the teaching that is objective and unbiased. Although each observation included in this study was conducted by the researcher, a range of test evaluators used the instrument in a variety of contexts to test whether the outcome data was stable and consistent. On each of the test occasions, after only two calibrating sessions, observers developed consistency in their measurements and achieved scores within three points for the total score and no more than a two-point difference in each domain.

Validity is concerned with making sure the instrument is measuring what it is designed to measure, which is instructional quality. At very minimum, it appears from use that the PAQT has face validity, which means people generally believe the data and are confident that it provides valuable and credible information about instruction. Content validity, on the other hand, has more rigorous expectations (Merriam, 2002). The method to define and measure instruction described in this study is only one way among many others. Because the instrument was developed directly from practitioners and then confirmed and revised so that it was consistent with current research, content validity is strengthened. Confirmatory methods (Tashakkori & Teddlie, 1998) were utilized early in the implementation phase in the area of English/language arts. The researcher and two English content-area experts conducted joint observations of four English lessons using different methods. One expert used a version of the Framework for Teaching (Danielson,

2007), while the other used a district form focusing on differentiated instruction.

Although the instruments generated different kinds of data, the English experts agreed that the PAQT assessment yielded the most accurate and concise description of the instruction observed. For math, ten lessons were jointly observed by the researcher and a math expert. After four lessons, which were used for training and calibration, the math expert began using the PAQT and felt strongly that it yielded accurate and useful results. The inter-rater reliability was very strong. Of the six remaining lessons, it yielded the same score 50% of the time, was within two points 40% of the time, and was within three points 10% of the time.

Because each classroom was observed only once, reliability as it pertains to stability of teacher behavior or the internal consistency of the measurement scale cannot be reported. Additionally, criterion-related validity, or the extent to which the observational measures relate to an agreed-upon level of performance, cannot be reported in the absence of additional research. Establishing concurrent validity or the extent to which the instrument is related to other instruments is beyond the scope of this study.

Data Analysis

Multilevel linear modeling addressed the major methodological shortcomings of typical quantitative studies that fail to distinguish between classroom- and student-level effects (Heck, Thomas, & Tabata, 2010). This is often a significant methodological challenge for educational researchers and evaluators who are attempting to link program implementation factors typically measured at the classroom or teacher level with student outcomes measured at the individual level (Newton & Llosa, 2010; Raudenbush & Bryk, 2002). Multilevel models of student learning assume that students are not randomly

assigned to classrooms and teachers are not randomly distributed across schools (Heck, 2007). In reality, students with varied skills and abilities are clustered in classrooms with teachers of differing attitudes, skills, and demeanors. Disentangling the factors that make one teacher more instructionally effective with students having different skill sets requires methods that allow one to account for both student-level variables and the unique contributions made by teachers at the classroom level as they interact with students.

In the past, repeated observations (e.g., test scores) on individuals were primarily conducted using repeated measures analysis of variance (RM-ANOVA), but this type of framework has a few limitations. First, the analysis cannot consider the groupings of individuals within classrooms or teachers. Second, the approach requires complete data; therefore, because there is likely to be missing data present in longitudinal analyses, the RM-ANOVA can result in substantial losses of information. Third, the changes observed within individuals must be assumed to be “fixed”; that is, changes must be assumed to be the same for every individual in the study (Hox, 2010). More recently, longitudinal analyses of student learning have also been conceptualized as “random coefficients” or mixed models. The advantages of this latter approach are that the nesting of individuals within higher-level groups can be easily incorporated into the model, missing data on students’ test scores can be incorporated in the analyses, and individuals’ beginning achievement scores as well as changes (or growth) in their scores over time can be considered as randomly varying between students and higher groupings such as classrooms (Hox, 2010). These features can provide a more thorough examination of

student academic growth as well as the possible impact of teachers on academic changes in students.

In this framework, the repeated measures within individuals (RSID) are specified at level 1, and measures between individuals within classes (RSID*CLASSID) are specified at level 2. The nesting of individuals within classrooms (level 3) and teachers (level 4) are considered at successive levels of the analysis. In the sets of mixed models for the English and math dependent variables presented, there are two primary parameters of interest. The intercept in each model represents the grand mean of all 2011 CST scaled scores, or the starting points for each student. On the CST in both English and math, a score of 350 is considered a proficient score. The growth parameter in each model is defined as the change between the 2011 CST scaled score and the 2012 CST scaled score. Estimates from each of the four-level models were computed with the SPSS Statistics Version 20.0 (IBM Corp., 2011).

Model Specification

Question 1(a): How much do Student Growth Trajectories Vary?

Model 1: Baseline Model: Null Model/No Predictors

The first step in developing the multilevel model was to develop a null model with no predictors (independent variables). This is the same as conducting a one-way ANOVA (analysis of variance). The dual purposes of this step are: (a) to estimate the mean starting point (intercept) and the mean growth trajectory (*change11–12*) for all students, and (2) to partition the variance into its within- and between-group components (Heck, Thomas, & Tabata, 2010). If the between-group (students, classroom, or teacher) variance turns out to be significant, then a multilevel model is justified.

In a growth model, the baseline model has no predictors except for the variable representing the change between the CST scaled score in 2011 and the CST scaled score in 2012. This facilitates an initial assessment of the average change in scores between the two years. At the student level, the baseline model can be represented by the following equation:

$$Y_{ij} = \pi_{0ij} + \pi_{1ij}(\text{change11-12})_{ij} + \varepsilon_{ij}, \quad (1)$$

where Y_{ij} represents the achievement score at time (t) for individual student (i) in classroom (j). The intercept π_{0ij} is the grand mean parameter at the initial time ($t = 0$) of measurement (i.e., the 2011 CST score). The slope parameter $\pi_{1ij}(\text{change11-12})_{ij}$ represents the average linear growth rate from 2011 to 2012 of individual (i) in classroom (j) and serves as the key parameter in the study. The ε_{ij} symbol represents the error associated with measuring each individual's true change trajectory. As summarized in Eq. 1, this suggests that a score of zero for the growth parameter would represent no change between the two measurements (2011 and 2012) since the *change11-12* parameter represents the change in 2011 math or reading scores for 2012 (i.e., the end of the next higher grade). For example, if the 2011 intercept were 350, a growth parameter of +5.0 would indicate that on average student scores increased 5.0 scaled score points from 2011 to 2012. The 2012 intercept representing average achievement would then be estimated to be 355. In contrast, if the growth parameter were -5.0, this would indicate that the average student scores dropped by 5 scaled score points between years. In this case, then, the 2012 mean would be estimated as only 345.

At level 2 of the null model (between students), the model is specified as follows:

$$\pi_{0ij} = \beta_{0j} + r_{0ij}. \quad (2)$$

At this second level, β_{0j} represents the mean score at time zero (2011 CST) for all students, and r_{0ij} represents the variability (or residual) in estimating each individual's score. This suggests that initial levels of achievement can vary between individuals within their classrooms. This is an important point because it would be unlikely that a teacher would be assigned to a classroom with students who are all at the same level of initial achievement.

At level 3 (classes), the model is specified as follows:

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad (3)$$

where γ_{00} represents the intercept (grand mean 2011 score), and u_{0j} represents the random variation of each class's mean from the grand mean with the subscript j indicating that the intercept varies across classrooms.

Question 1(b): How Do Student Characteristics Affect Growth Trajectories?

Model 2: Adding Student Characteristics

Prior to making inferences about growth associated with instructional quality, the effects of exogenous student variables that have been shown to account for variance in student achievement were factored into the analysis. According to Salganik (1994), covariates that are used to adjust teacher effectiveness scores should be related to student achievement, beyond the control of the school to change, and accepted as legitimately associated with challenges facing the school (Newton, et al., 2010; Salganik, 1994).

These factors include students' prior year's achievement, school attendance, student behavior, socioeconomic status, ethnicity, and program participation (e.g., English Learner Program [EL], Special Education [IEP], and gifted and talented education [GATE]).

After significant variance in student growth trajectories was established in both mathematics and English, a set of student-level predictors was added to the model to explain the differences in students' initial achievement and their growth trajectories. The model remains the same at level 1 (see Eq. 1). At level 2 (between students), model 2 is represented by the equation:

$$\pi_{0ij} = \beta_{0j} + \beta_{1j}EL_{ij} + \beta_{2j}AA_{ij} + \dots + \beta_{7j}Zattend_{ij} + r_{0ij}. \quad (4)$$

The first coefficient (β_{0j}) represents the initial achievement for all students. The successive regression coefficients ($\beta_{1j} - \beta_{7j}$) account for different student backgrounds, participation in special programs (e.g., EL, IEP, and GATE), as well as behavioral and attendance patterns. The latter two variables in the models are standardized with a mean of zero and a standard deviation of one to improve the interpretability of the intercepts describing the student outcomes. Once again, r_{0ij} represents the random intercept parameter describing variance in estimating parameters at the student level. For this model, the classroom level remains as in Eq. 3.

Model 3: Adding a Random Slope for Change at the Classroom Level

In model 3, the only change was to add a random slope at the student and classroom levels to see if student growth trajectories vary across students and classrooms.

In this model, all of the previous equations remain the same. The random growth slope that captures variance between students at level 2 is represented by the equation:

$$\pi_{1ij} = \beta_{10j} + r_{1ij} \quad (5)$$

At the classroom level (level 3), a random growth slope is also added that estimates the variance in the student growth (*change11-12*) as follows:

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (6)$$

Eq. 6 therefore implies that students achieve differing levels of growth at the classroom level during the 2012 academic year. This random parameter at the classroom level becomes the key parameter in determining how much “added growth” teachers of varying quality may add to student progress.

Question 1(c): Do Classroom and Teacher Characteristics Affect Student Growth Trajectories?

Model 4: Adding Classroom-Level Predictors.

In model 4, the classroom predictors are added to the third level of the hierarchy. The level 1 model remains the same (see Eq. 1), as does the level 2 model (see Eq. 4–6). At the classroom level (level 3), the mean socio-economic status (SES) score (ZSESmean) and the instructional quality (ZPAQT) scores were added into the model, as represented by the following two equations:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}ZSESmean_j + \gamma_{02}ZPAQT_j + u_{0j} \quad (7)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}ZPAQT_j + u_{1j} \quad (8)$$

Both predictors were also standardized with a mean of zero and a standard deviation of one and added to the intercept model. This allows us to view the classroom intercepts as the level of achievement in a classroom of “average” SES composition and “average” instructional quality. Four random effects (two random intercepts and two random slopes for *change11–12* at the student and classroom levels) are represented as r_{0j} and r_{1j} (at level 2) and u_{0j} and u_{1j} (at level 3). In addition, the effect of instructional quality is entered in the model as an interaction (*change11–12*ZPAQT*) as the result of substituting the level 3 models (Eq. 6) into the level 2 and level 1 models.

Model 5: Adding Teachers as a Fourth Level in the Hierarchy

A final model in math and English added the teacher level, since teachers could have taught one or more classes. No predictors were added at the teacher level to explain differences between teachers; however, the final models allow deeper investigation of the differential instructional quality effect on student growth in achievement during the year of the study. To accommodate the teacher level, the previous equations are altered by adding a k subscript to designate teachers. To illustrate, the level 1 becomes:

$$Y_{ijk} = \pi_{0ijk} + \pi_{1ijk} \text{change}_{ijk} + \varepsilon_{ijk} \quad (9)$$

The level 2 and 3 models stay the same except for adding the k subscripts. At level 4, no teacher-level predictors are added; however, the random effects for teachers are specified as v_{00k} and v_{10k} . The equations are as follows:

$$\gamma_{00k} = \nu_{000} + v_{00k} \quad (10)$$

$$\gamma_{1k} = \nu_{100} + v_{10k} \quad (11)$$

For equations that include parameters that are converted to z-scores, the original standard deviations are listed in Table 5.

Table 5

Standard deviations for parameters converted to z-scores

	<i>N</i>	Minimum	Maximum	Mean	Std. Deviation
PAQT	2949	36.0	98.0	65.24	14.66
SESmean	2949	12.1	91.3	61.31	17.09
Attendance	2949	0	63.4	6.48	7.3
Behavior	2949	0	10	.17	.719

Question 2: What instructional practices yield the highest and lowest levels of instructional quality?

The final stage of analysis represents a significant step toward making public the characteristics of high- and low-quality instruction. The information generated from the comparative analysis of instructional tasks has the potential to inform professional development decisions for both teachers and administrators. The study's design will have been effective only if it causes educators to reflect on the quality of the instruction students receive and continually develop their ability to facilitate their own learning so that their practices result in high-quality instruction for all students.

In the spirit of fostering growth and change, the second research question in the study focuses on elucidating the teacher actions that yielded the highest and lowest levels of observed student learning and student achievement. High- and low-quality classrooms were identified by creating a matrix to show which classrooms with high levels of instruction yielded growth in student achievement, and which classrooms with low levels

of instruction yielded losses in student achievement. The rationale for this model is based on the fact that, in accordance with the No Child Left Behind Act (NCLB, 2001), schools are judged by the percent of students who score proficient or advanced in English/language arts and mathematics. Therefore, it is useful to compare the mean growth of students in classrooms to ascertain the aggregate effect of instruction on achievement while other factors are taken into account. In this model, a gain of zero is considered a positive outcome because scaled scores represent the degree of proficiency with the grade-level content and the complexity of the content advances as the grades progress. As such, maintaining the same scaled score represents a gain in the mastery of content. For example, if a student earned a scaled score in mathematics of 379 at the end of 6th grade and a scaled score of 379 at the end of 7th grade, then that student maintained his or her proficiency despite having been tested on more challenging content.

Instructional quality in this study is dependent on both instruction and achievement, so to answer question two, I selected for further analysis those cases where instructional scores (PAQT) were above the mean and class aggregates of student achievement gains were zero or above (Growth). These high-quality lessons were analyzed to understand the teacher actions that promoted high levels of student engagement and high levels of learning as measured by growth on the CST. In contrast, the descriptive data from lesson observations where the instructional score (PAQT) fell below the mean and aggregate student achievement was less than zero were analyzed to understand why certain practices inhibit or prevent student engagement and student learning.

Summary of Methodology

This chapter described the methodology employed in this parallel-convergent mixed-methods study of instructional quality. The goal was to use multilevel modeling to analyze instructional data at the classroom and teacher levels to determine whether the quality of the instruction students received impacted their growth trajectories in English/language arts and mathematics. The sample included 6th, 7th, and 8th grade students from two California middle schools, and the data included students' scaled scores for two consecutive years on the California Standards Test (CST) as well as classroom-level data collected through observations of instruction. In this chapter, I explained procedures for addressing issues of missing data, and I described in detail the observation instrument to demonstrate its potential for providing informative data to teachers regarding the impact of their instructional decisions on student learning. Additionally, I discussed the reliability and validity of the instrument. Finally, I provided a description of the multilevel analysis and qualitative cross-case narrative analysis corresponding to each research question.

CHAPTER FOUR: RESULTS

This chapter presents the results of the study. The first section of the chapter presents the results for each level of quantitative analysis beginning at the student level and moving to the classroom and teacher levels for mathematics. Next, the quantitative results for English/language arts are presented. The chapter concludes with the qualitative analysis, which provides descriptions of eight quality practices that were

observed by the highest performing teachers and six ineffective practices that were consistently observed in low-quality classrooms.

Mathematics

Model 1: Null Model/No Predictors

Table 6 provides the results of the baseline mathematics model, which summarizes the average beginning achievement levels of students (at the end of the 2011 academic year) and the change in their math scores over the 2012 school year. The intercept represents the average 2011 classroom achievement score for all students (374.59, $p < .001$). The growth parameter (*change11–12*) shows the change in student test scores during the 2012 school year. In general, this parameter indicates that average student achievement in math decreased by almost 19 scaled score points from 2011 to 2012 ($\gamma_{10} = -18.79$, $p < .001$).

Table 6

<i>Estimates of mathematics fixed effects^a for model 1</i>					
Parameter	Estimate	Std. Error	df	T	Sig.
Intercept	374.59	6.77	54.59	55.29	0.00
Change11–12	-18.72	1.40	1426.97	-13.40	0.00

Note: a. Dependent Variable: mathscore.

Table 7 presents the variance decomposition in initial achievement for the baseline model. On this and subsequent models, I follow Hox's (2010) recommendation that, since variance estimates cannot be below zero, a one-tailed test is sufficient for testing variance components in multilevel models. This means that significance levels are

divided by two when reporting variance components. The residual variance, which represents differences in initial math achievement within individual students (RSID at level 1), is 1347.22 ($p < .001$). This represents 23.5% of the total variance in initial math achievement. The variance in initial achievement (RSID*Class_ID) between individual students (level 2) is 2038.76 (one-tailed $p < .001$), which represents 35.6% of the total variance shown in this model. The variance associated with classrooms (Class_ID) is 2344.41 (one-tailed $p < .001$). This represents 40.9% of the total variance in initial math achievement. The results of this preliminary analysis therefore suggest that, after accounting for individual differences in achievement, the variance between classrooms at level 3 is still highly significant, which justifies looking deeper into what might account for those differences.

Table 7

<i>Estimates of mathematics covariance parameters^a for model 1</i>					
Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		1347.22	51.8	26.01	.000
Intercept [Class_ID]	Variance	2344.41	474.29	4.94	.000
Intercept [RSID * Class_ID]	Variance	2038.76	104.45	19.52	.000

Note: a. Dependent Variable: mathscore.

Model 2: Adding Student Characteristics

In model 2, student level characteristics were introduced. Other factors that likely influence achievement include student behavior, attendance, ethnicity, and whether or not students participate in special programs like special education (IEP), the English Learner

program (EL), or gifted and talented education (GATE). These factors are added into the model as covariates and represent the exogenous variables that are outside the teachers' control but may affect student growth trajectories both for individual students (for example, behavior or attendance) or groups of students (for example, English Learners). If prior achievement (scores in 2011) and other factors are accounted for, and there still remains significant variance between classrooms, then level three of the model is justified.

For math, the estimates of fixed effects are summarized in Table 8. After adjusting for student-level variables, the initial classroom achievement intercept increased slightly from 374.59 to 377.61, and the average growth changed from -18.72 to -18.32. Each of the parameter estimates corresponds with the level of adjustment to student test scores (2012 CST standard score) associated with a unit increase in the specific student-level predictor. For example, compared with students not receiving EL services, students receiving EL services had a predicted initial math score approximately 25.81 points lower than the intercept. Compared with students having "average" attendance, a student with a one standard-deviation increase in attendance problems (i.e., approximately 7 days absent) would have a predicted initial achievement score of 11.92 points lower than the intercept. All of the covariates used to account for individual differences among students significantly affect achievement levels ($p < .001$).

Table 8

Estimates of mathematics fixed effects^a for model 2

Parameter	Estimate	Std. Error	df	T	Sig.
Intercept	377.61	5.15	70.89	73.33	0.00
EL	-25.81	3.57	1591.15	-7.23	0.00

AA	-19.40	3.93	1561.70	-4.94	0.00
HISP	-17.70	2.79	1532.97	-6.34	0.00
Gate	33.88	3.03	1528.26	11.18	0.00
IEP	-18.06	5.53	1690.08	-3.27	0.00
ZBehavior	-4.81	1.49	1715.71	-3.23	0.00
ZAttendance	-11.93	1.54	1589.27	-7.75	0.00
Change11-12	-18.32	1.39	1447.02	-13.17	0.00

Note: a. Dependent Variable: mathscore.

Table 9 presents the variance decomposition for model 2. Adding student-level factors changed the way that the variance is distributed and decreased the variance at the student level by about 27.4% (from 2038.76 to 1479.48) and at the classroom level by about half (51.5%; from 2344.41 to 1138.02). This finding indicates that student-level factors do impact student performance; however, even after adjusting classrooms for student factors, there is still a substantial portion of variance (28.7%) attributed to the classroom level, which serves as our justification for adding an additional model to explain this variability further.

Table 9

<i>Estimates of mathematics covariance parameters^a for model 2</i>					
Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		1343.26	51.47	26.10	0.00
Intercept [Class_ID]	Variance	1138.02	240.23	4.74	0.00

Intercept	Variance	1479.48	84.86	17.43	0.00
[RSID*Class_ID]					

Note: a. Dependent Variable: mathscore.

Model 3: Adding a Random Slope for Change at the Classroom Level

Model 3 looks at the differences in student growth trajectories at the classroom level by allowing the individual slopes to vary across classrooms. The random slope and intercept model assumes that all of the individual change trajectories have the same algebraic form, but not every individual has the same trajectory. Because each person has a different intercept (initial achievement) and slope coefficient (*change11–12*), the random slope formulation implies that student growth from 2011 to 2012 varied according to their classroom placements.

Table 10 shows that the fixed effects are all still highly significant and have remained stable. There is a slight increase in the negative effects of having an IEP and a slight decrease in the negative direction of student growth during 2012.

Table 10

Estimates of mathematics fixed effects^a for model 3

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	376.91	5.80	64.17	64.94	0.00
EL	-26.11	3.57	1578.36	-7.32	0.00
AA	-18.99	3.92	1555.79	-4.85	0.00
HISP	-17.67	2.79	1530.92	-6.34	0.00
Gate	33.60	3.04	1525.65	11.05	0.00

IEP	-19.69	5.51	1670.22	-3.57	0.00
ZBehavior	-5.04	1.48	1687.63	-3.41	0.00
ZAttendance	-11.68	1.54	1579.98	-7.61	0.00
Change11–12	-16.89	3.52	52.93	-4.80	0.00

Note: a. Dependent Variable: mathscore.

More importantly, model 3 adds another dimension to the covariance parameters. The covariance parameters presented in Table 11 suggest that student growth during 2012 varies across classrooms (variance = 578.64, Wald $Z = 4.460$, one-tailed $p < .001$). This finding suggests that the classroom level has a significant impact on student growth trajectories. This variation may reflect the way students are grouped for math classes; however, it may also represent the impact of differential instructional quality. It should be noted that adding the random slope parameter at the classroom level changes the estimation of the variance components for the initial intercept parameters.

Table 11

Estimates of mathematics covariance parameters^a for model 3

Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		1044.55	41.06	25.44	0.00
Intercept [Class_ID]	Variance	1529.00	320.79	4.77	0.00
Change11–12 [Class_ID]	Variance	578.64	129.74	4.46	0.00

Intercept					
	Variance	1636.37	83.32	19.64	0.00
[RSID*Class_ID]					

Note: a. Dependent Variable: mathscore.

Model 4: Adding Classroom-Level Characteristics

The original hypothesis for this study was that instructional quality plays an important role in determining student achievement trajectories. Model 4 adds two classroom variables in an attempt to explain some of the variance in math achievement between classrooms. The first is the PAQT score, which was derived from classroom observations that were conducted during the school year. The second is the aggregate socio-economic status (SESmean) for each class because research has shown that SES impacts achievement scores (Newton, et al., 2010).

The fixed effects in Table 12 show the coefficients for each factor. First, for the initial 2011 math score, the standardized PAQT score is unrelated to initial math achievement ($\gamma_{02} = -5.70, p = .207$). As expected, this finding indicates that initial student achievement in math was unrelated to the instructional quality scores of teachers because the teachers had not taught these particular students in 2011. In contrast, the classroom level SES statistic was significant ($\gamma_{01} = -23.41, p < .001$), suggesting that student test scores are impacted by average class SES levels. Specifically, for every one standard deviation (approximately 17 points) below the mean (61.3% free/reduced lunch), a student's test score will decrease by 23.4 scaled score points. Second, the interaction between student growth and instructional quality (PAQT) was both significant and

substantial ($\gamma_{11} = 8.85, p = .01$). The results suggest that for each standard deviation (approximately 15 points) increase in teacher quality (above the mean of 65), student achievement scores increase by almost 9 scaled score points, holding all other variables constant.

Table 12

Estimates of mathematics fixed effects^a for model 4

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	377.07	4.77	68.89	79.08	0.00
ZSESmean	-23.41	4.50	51.27	-5.20	0.00
ZPAQT	-5.70	4.45	49.84	-1.28	0.21
EL	-26.13	3.56	1586.57	-7.34	0.00
AA	-18.58	3.92	1556.56	-4.74	0.00
HISP	-17.50	2.79	1532.22	-6.27	0.00
Gate	32.88	3.05	1518.84	10.80	0.00
IEP	-19.94	5.50	1680.13	-3.62	0.00
ZBehavior	-4.99	1.48	1692.09	-3.37	0.00
ZAttendance	-11.74	1.53	1584.54	-7.65	0.00
Change11–12	-16.87	3.31	52.37	-5.10	0.00
Change11–12 *	8.85	3.29	51.52	2.69	0.01
ZPAQT					

Note: a. Dependent Variable: mathscore.

Table 13 presents the variance decomposition for model 4. The effect of the random slope at the classroom level again impacts the distribution of variance. The between-classroom variance in random slopes decreased from 578.64 (Table 11) to 498.87 (Table 13), which is about a 13.8% reduction. This decrease in slope variance at the classroom level suggests that instructional quality (ZPAQT) accounted for significant reductions in variability in students' math achievement at the classroom level. The reduction in classroom variance combined with the fixed effect of the interaction between instructional quality and student growth trajectories (ZPAQT**change11–12*) suggest that teachers with higher PAQT scores produced considerably higher student math scores during the 2012 school year.

Table 13

<i>Estimates of mathematics covariance parameters^a for model 4</i>					
Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		1047.10	41.24	25.39	0.00
Intercept [Class_ID]	Variance	972.75	215.46	4.52	0.00
Change11–12 [Class_ID]	Variance	498.87	114.80	4.35	0.00
Intercept [RSID*Class_ID]	Variance	1647.23	83.88	19.64	0.00

Note: a. Dependent Variable: mathscore.

Model 5: Adding Teachers as a Fourth Level in the Hierarchy

In the final model, teachers were added as a fourth level in the analysis. Because all of the teachers in the study taught more than one period of math, both the instructional

score (PAQT) and their student achievement growth could be aggregated across all classes and students. Table 14 summarizes the fixed effects. All of the parameters remain substantial and significant, but adding the teacher level increases the size of the PAQT effect from 8.85 to 11.01. The stronger effect on student “change” due to the ZPAQT variable ($U_{100}=11.01, p = .012$) may be an indicator of the impact of instructional quality at the teacher level on student math scores.

Table 14

Estimates of mathematics fixed effects^a for model 5

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	379.66	6.81	10.47	55.79	0.00
ZSESmean	-27.66	4.95	47.73	-5.58	0.00
ZPAQT	-4.05	5.65	26.64	-0.72	0.48
EL	-26.25	3.56	1586.91	-7.37	0.00
AA	-18.60	3.92	1557.63	-4.75	0.00
HISP	-17.44	2.79	1532.78	-6.26	0.00
Gate	32.95	3.04	1520.58	10.83	0.00
IEP	-19.87	5.50	1679.67	-3.61	0.00
ZBehavior	-5.01	1.48	1689.72	-3.39	0.00
ZAttendance	-11.77	1.53	1585.98	-7.67	0.00
Change11–12	-19.94	5.61	11.88	-3.56	0.00
Change11–12 *	11.01	4.20	44.62	2.62	0.01
ZPAQT					

Note: a. Dependent Variable: mathscore.

Adding teachers to the model adds two additional covariance parameters as shown in Table 15. The first represents the variance in initial achievement accounted for by the teacher level, and the second represents the variance associated with the interaction between student growth trajectories (*change11–12*) and teachers. We see that the initial intercept does not vary across teachers (Wald $Z = 1.37$, one-tailed $p = 0.085$), which implies that teachers were teaching students who were relatively “similar” in achievement variability in 2011.

Importantly, the variability in student growth due to teacher level was significant (Wald $Z = 1.913$, one-tailed $p = .028$), suggesting that teachers produce differing amounts of student learning in math. Similarly, the table suggests variability in math growth is also significant across classrooms (variance = 226.00, Wald $Z = 3.306$, one-tailed $p < .001$). These results imply that student growth in achievement is related to both their classroom placement and their teachers and likely related to differences in instructional quality.

Table 15

<i>Estimates of mathematics covariance parameters^a for model 5</i>					
Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		1046.98	41.24	25.39	0.00
Intercept					
	Variance	403.66	294.38	1.37	0.17
[Teacher_ID]					
Change11–12					
	Variance	344.39	180.06	1.91	0.06
[Teacher_ID]					

Intercept [RClass_I * Teacher_ID]	Variance	664.68	181.12	3.67	0.00
Change11–12					
[RClass_I * Teacher_ID]	Variance	226.00	68.35	3.31	0.00
Intercept [RSID * RClass_I * Teacher_ID]	Variance	1647.42	83.88	19.64	0.00

Note: a. Dependent Variable: mathscore.

English/Language Arts

The next section provides the same series of models to examine students' English outcomes over the length of the study. In general, these results suggest that, although the effects of the instructional quality variable (ZPAQT) were positive and statistically significant, they were not as substantial as in math.

Model 1: Null Model/No Predictors

In Table 16, the baseline model for English indicated the intercept (mean for 2011 scores) was 363.47. In contrast to math, the data in Table 16 indicates that English scores on average increased slightly by 0.80 over the course of the study.

Table 16

Estimates of mathematics fixed effects^a for model 1

Parameter	Estimate	Std. Error	df	T	Sig.
Intercept	363.47	4.03	59.43	90.20	0.00

Change11-12	0.80	0.85	1566.21	0.94	0.35
-------------	------	------	---------	------	------

Note: a. Dependent Variable: Englishscore.

The variance components show that most of the variance in initial English scores was between individuals (variance = 1656.22, one-tailed $p < .001$), which accounts for 53.6%. Variance at the classroom level accounted for 28.2%, and variance within students accounted for 18.1%. Table 17 presents the variance decomposition in initial achievement for the baseline model. The residual variance represents differences in initial English achievement within individual students (level 1), which was 559.78 ($p < .001$). This represents 23.5% of the total variance in initial English achievement. The results of this preliminary analysis therefore suggest that, after accounting for individual differences in achievement, the variance between classrooms at level 3 is still highly significant, which justifies further analysis to account for those differences.

Table 17

Estimates of English covariance parameters^a for model 1

Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		559.78	20.35	27.50	0.00
Intercept [Class_ID * Teacher_ID]	Variance	871.92	175.62	4.97	0.00
Intercept [RSID * Class_ID]	Variance	1656.22	68.44	24.20	0.00

Note: a. Dependent Variable: Englishscore.

Model 2: Adding Student Characteristics

Table 18 indicates that adding student-level characteristics substantially changes the intercept from 363.47 in the baseline model to 373.78 in model 2. The negative effects of being an English Learner, having an IEP, or being African American were significantly more pronounced in the English models than they were in the math models. Behavior had a stronger negative effect while attendance had less of a negative effect in English than in math. The growth parameter (*change11–12*) suggests that, on average, students' scores remained about the same after adjustment for the background variables in the model.

Table 18

<i>Estimates of English fixed effects^a for model 2</i>					
Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	373.78	2.67	109.16	139.87	0.00
EL	-38.20	2.72	1786.92	-14.07	0.00
AA	-24.56	2.88	1739.41	-8.54	0.00
HISP	-18.37	2.09	1716.59	-8.80	0.00
Gate	23.60	2.33	1654.19	10.15	0.00
IEP	-30.40	4.61	1820.38	-6.59	0.00
ZBehavior	-7.30	1.14	1852.65	-6.37	0.00
ZAttendance	-6.13	1.13	1803.01	-5.45	0.00
Change11–12	0.82	0.85	1587.21	0.97	0.33

Note: a. Dependent Variable: Englishscore.

Table 19 suggests that adding the student-level factors reduced variability in initial achievement between students substantially from 1656.22 to 1114.52 (32.7%). Classroom-level variance in initial achievement was dramatically reduced from 871.92 to 242.01 (72.2%). Importantly, this implies at the classroom level that differences in students' initial English scores are largely due to differences in classroom composition.

Table 19

<i>Estimates of English covariance parameters^a for Model 2</i>					
Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		558.80	20.27	27.57	0.00
Intercept [Class_ID]	Variance	242.01	56.49	4.28	0.00
Intercept [RSID * Class_ID]	Variance	1114.51	50.18	22.21	0.00

Note: a. Dependent Variable: Englishscore.

Model 3: Adding a Random Slope for Change at the Classroom Level

As seen in math, adding the random slope component for levels 2 (students) and 3 (classrooms) does not significantly impact the fixed effects (see Table 20). As stated earlier, this model assumes that all of the individual change trajectories have the same algebraic form, but not every individual has the same trajectory. Except for the change variable, the fixed effects are all still highly significant and have remained stable. The less significant change variable may reflect the lack of change overall in achievement for students in English.

Table 20

Estimates of English fixed effects^a for Model 3

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	374.07	2.62	106.06	142.66	0.00
EL	-38.36	2.71	1792.15	-14.16	0.00
AA	-24.70	2.87	1741.05	-8.61	0.00
HISP	-18.39	2.08	1716.96	-8.84	0.00
Gate	23.76	2.32	1646.37	10.26	0.00
IEP	-29.99	4.58	1795.44	-6.55	0.00
ZBehavior	-7.46	1.14	1853.45	-6.53	0.00
ZAttendance	-5.91	1.12	1777.35	-5.29	0.00
Change11–12	0.35	1.26	59.41	0.28	0.78

Note: a. Dependent Variable: Englishscore.

The variance components, on the other hand, are affected when we add the random change slope between students and classrooms. Table 21 suggests that student change in English is related to differences between students (variance = 199.46, $p < .001$) and differences between classrooms (variance = 49.94, one-tailed $p = .0015$). The next model will attempt to explain some of those differences.

Table 21

Estimates of English covariance parameters^a for model 3

Parameter	Estimate	Std. Error	Wald Z	Sig.
Residual	433.29	35.17	12.32	0.00
Intercept [Class_ID] Variance	230.37	55.11	4.18	0.00

Change11–12 [Class_ID]	Variance	49.94	16.69	2.99	0.00
Intercept [RSID * Class_ID]	Variance	1123.33	49.87	22.53	0.00
Change11–12 [RSID * Class_ID]	Variance	199.45	62.38	3.20	0.00

Note: a. Dependent Variable: Englishscore.

Model 4: Adding Classroom-Level Characteristics

Table 22 estimates the fixed effects for model 4 when classroom-level characteristics are added into the model. Like in math, the SES mean (ZSESmean) was significant but not as substantially negative (-12.6 compared to -23.5 in math). However, the negative effects for English Learners, African American students, and students with IEPs were still quite substantial and highly significant. The interaction between student growth trajectories and instructional quality (*change11–12*ZPAQT*) was less substantial than in math, but still positive and significant ($\gamma_{11} = 2.57, p. < .05$).

Table 22

Estimates of English fixed effects^a for model 4

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	374.70	2.16	152.98	172.94	0.00
ZSESmean	-12.52	1.79	58.81	-7.03	0.00
ZPAQT	-1.04	1.81	51.74	-0.67	0.51
EL	-37.68	2.71	1781.42	-14.29	0.00

AA	-24.08	2.89	1749.22	-8.54	0.00
HISP	-18.16	2.10	1724.64	-8.54	0.00
Gate	22.96	2.34	1633.31	9.81	0.00
IEP	-30.15	4.58	1834.08	-6.59	0.00
ZBehavior	-7.49	1.15	1862.67	-6.51	0.00
ZAttendance	-6.55	1.12	1791.18	-5.83	0.00
Change11–12	0.56	1.22	59.36	0.45	0.65
Change11–12 *	2.57	1.27	52.93	2.03	0.05
ZPAQT					

Note: a. Dependent Variable: Englishscore.

The variance components in Table 23 suggest that the *change11–12* variable varies across classrooms and students. The significant ZPAQT**change11–12* interaction at the classroom level (variance = 44.68, Wald Z = 2.85, one-tailed $p < .002$) implies that teachers are partly responsible for the variance in student change from 2011 to 2012. The *change11–12* variance (variance = 195.46, one-tailed $p < .001$) shows that changes in scores also vary between students, meaning there are significant differences in student learning in English at both the student level within classrooms and between classrooms.

Table 23

Estimates of English covariance parameters^a for model 4

Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		435.74	35.66	12.22	0.00
Intercept [Class_ID]	Variance	99.44	30.31	3.28	0.00

Change11–12 [Class_ID]	Variance	44.69	15.68	2.85	0.00
Intercept [RSID * Class_ID]	Variance	1154.84	50.99	22.65	0.00
Change11–12 [RSID * Class_ID]	Variance	195.46	63.28	3.09	0.00

Note: a. Dependent Variable: Englishscore.

Model 5: Adding Teachers as a Fourth Level in the Hierarchy

Teachers were added to the final model, creating a fourth level. The fixed effects in Table 24 remain similar to Model 4. Importantly, the interaction between changes in student growth trajectories and instructional quality at the teacher level gain in terms of the size of effect ($\nu_{100} = 3.67$, $p = .019$). At the teacher level, the results suggest that for every 1 standard deviation increase in teacher quality (15 points), the predicted student test scores in English would increase by 3.67 scaled score points.

Table 24

Estimates of English fixed effects^a for model 5

Parameter	Estimate	Std. Error	df	T	Sig.
Intercept	374.44	2.61	30.51	143.31	0.00
ZSESmean	-12.32	1.87	55.55	-6.58	0.00
ZPAQT	-0.06	2.04	39.70	-0.03	0.98
EL	-38.80	2.71	1772.08	-14.32	0.00
AA	-24.49	2.90	1750.28	-8.46	0.00

HISP	-17.84	2.10	1728.33	-8.49	0.00
Gate	22.38	2.35	1646.19	9.51	0.00
IEP	-29.84	4.54	1805.35	-6.57	0.00
ZBehavior	-7.31	1.15	1863.47	-6.35	0.00
ZAttendance	-6.72	1.13	1815.30	-5.95	0.00
Change11–12	0.44	1.61	11.26	0.27	0.79
Change11–12 *	3.67	1.47	29.35	2.49	0.02
ZPAQT					

Note: a. Dependent Variable: Englishscore.

The variance components presented in Table 25 show that there is no significant variance left across teachers in the random slope parameter (*change11–12*) after the other variables are accounted for. The change effect is not significant (Wald $Z = 1.13$, one-tailed $p = 0.129$); however, the effect of *change11–12* at the classroom level was significant (Wald $Z = 1.68$, one-tailed $p = .047$). Once again, the results from models 4 and 5 indicate that differences in instructional quality are related to differences in students' English outcomes at both the classroom and teacher levels.

Table 25

Estimates of English covariance parameters^a for Model 5

Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		533.91	19.71	27.09	0.00
Intercept					
	Variance	45.92	30.25	1.52	0.13
[Teacher_ID]					
Change11–12	Variance	23.32	20.62	1.13	0.26

[Teacher_ID]					
Intercept [RClass_I *	Variance	57.29	25.32	2.26	0.02
Teacher_ID]					
Change11–12					
[RClass_I *	Variance	26.74	15.95	1.68	0.09
Teacher_ID]					
Intercept [RSID *					
RClass_I *	Variance	1158.51	51.20	22.63	0.00
Teacher_ID]					

Note: a. Dependent Variable: Englishscore.

A Summary of the Quantitative Study

For mathematics, the final analytic sample included 1583 students from 54 classrooms with a total of 15 teachers. In English, 1773 students participated from 59 classrooms with a total of 18 teachers. The models at the fifth level show estimates of student's CST scores while taking into account many of the variables that are known to impact student achievement. All continuous variables were standardized (Attendance, Behavior, Class SES, and PAQT score). Variables for ethnicity and program participation were indicator variables, coded 1 = yes and 0 = no. The coefficients for the indicator variables represented the increment added to the intercept, i.e., average CST performance for a student's specific ethnicity and program. The coefficients for the continuous variables represented the increment added to the intercept for an increase of 1 SD in the independent variable (i.e., PAQT score and SES). A comparison of the tables reveals that the classroom SES mean had more than twice the effect in math as it did in

English. For example, in English (Table 24), if the class was one standard deviation below the mean in SES (17 points), the students' scores would be 12 points lower in English whereas they would be 27.7 points lower in math (Table 14). Conversely, English Learners were more negatively impacted in English (-38.8 points) than they were in math (-27.2). African American students earned on average 24.5 points less in English and 18.6 points less in math than their non-African American counterparts. The effects for Hispanic students were about equal in both subject areas (i.e., -17 points on average). GATE students scored on average 22 points higher in English and almost 33 points higher in mathematics, and IEP students scored 30 points lower in English and 20 points lower in math. Behavior was measured by the number of student suspensions. The data show that each suspension reduced a student's score by 7 points in English and 5 points in math. In contrast, attendance impacted students' scores almost twice as much in math as in English. For every 7 days that a student was absent, the student's score was reduced by 11.8 points in math and 6.7 points in English. Finally, the data show that for every 15 points a teacher scores above the mean (65) on the PAQT assessment, students gained 3.7 points in English and 11 points in math while all other factors were held constant. The analysis, using all predictors, explained about 25% of the variance among mathematics teachers and about 28% of the variance among English teachers in student performance.

Cross-Case Narrative Analysis of Instructional Practices

In the final section of this chapter, the qualitative analysis is presented to address the second research question about the instructional practices that yield the highest and lowest levels of instructional quality. In order to distinguish between high- and low-quality instruction, a matrix was created that combined student achievement data

(change11–12) with instructional quality (PAQT) data. This section shows that when student achievement and instructional quality are considered together, they serve as powerful indicators of effective teaching.

The resulting diagram partitions classrooms into four quadrants. The first quadrant, where most of the classrooms reside, represents classes that had low instruction and low achievement. Quadrant two represents classrooms that had low instruction, but high achievement. Quadrant three represents classrooms that had high instruction and low achievement. Finally, quadrant four represents classrooms with both high instruction and high achievement.

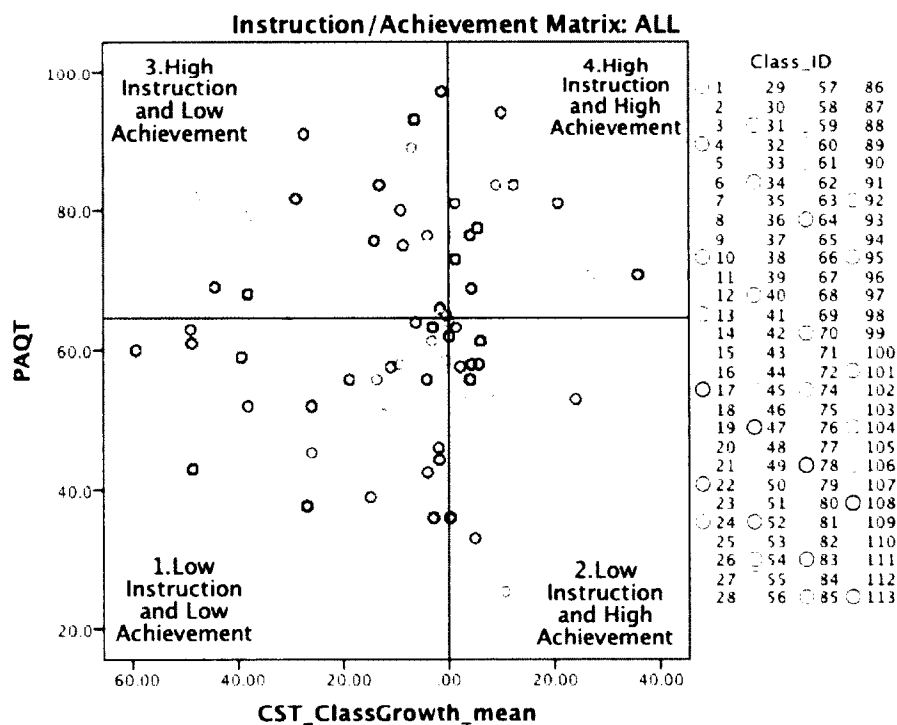


Figure 4. Instruction/achievement matrix: all classrooms///

In order to conduct the cross-case narrative analysis, the matrix was further divided by subject area. Quadrant four (upper-right) in both graphs represents the high-

quality classrooms and Quadrant one (lower-left) in both graphs represents the low quality classrooms.

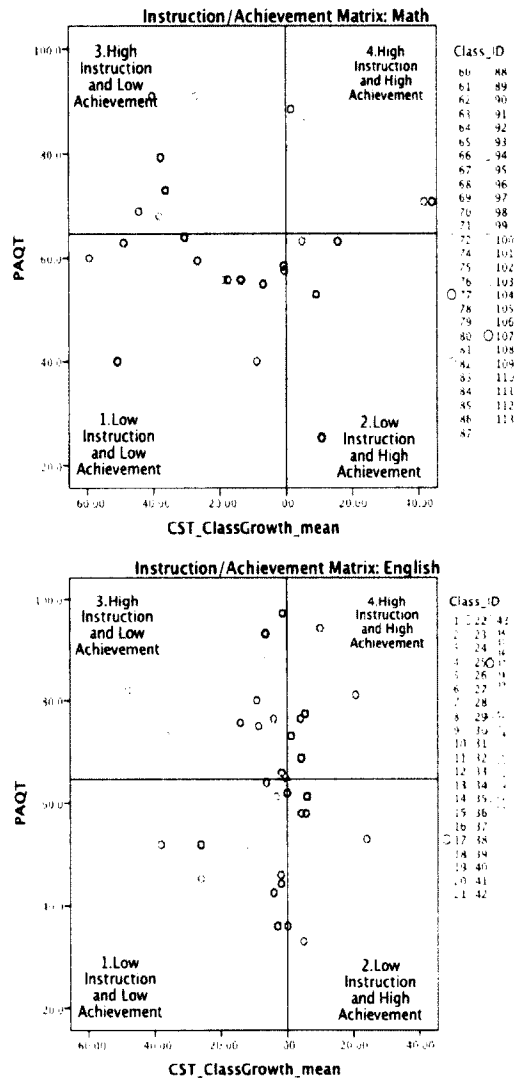


Figure 5. Instruction/achievement matrices for math and English

The descriptive analysis of classroom instruction revealed eight instructional practices that were implemented differently depending on the quality of instruction (PAQT) score. What distinguished high from low quality instruction was not necessarily whether or not the strategy was used, but how it was used to either engage or disengage students in learning. The practices identified include: (a) immediate engagement, (b)

scaffolded academic dialogue, (c) real-world connections, (d) front loading, (e), differentiated instruction, (f) feedback and conferring, (g) structured reflection, and (h) lesson closure with connected homework. The most artful teachers integrated the practices based on the feedback they received from their students as they taught their lessons. Some teachers used more than one practice at a time to create learning tasks that maximized students' participation, critical thinking, and students' use of academic language.

Many of the teachers in the low quality classrooms attempted some of the practices, but the way they implemented them prevented teachers from adding value to the interaction. In other cases, the practices were completely absent from the instruction. Alternatively, there were characteristics about the instruction in the low-quality classrooms that were absent from the high-quality classrooms. Ineffective practices are delineated in the final section of this chapter.

Immediate Engagement

A notable pattern amongst teachers in the highest quadrant was that they began teaching as soon as they had contact with the students and students knew exactly what to do upon entering the classroom. There was no wasted time or downtime, and students enjoyed the challenge of beginning their work right away. Usually the activity was a follow-up to some previous work, an assignment that was explained the last time they were together, or a warm-up activity that primed students' minds to engage in the next task.

High Quality: High Instruction/High Achievement

A common practice in mathematics was to have a warm-up problem posted on the board that either reviewed a concept that was taught previously or served as a pre-assessment for the work the class was going to engage in that day. Prior to a lesson on the order of operations, students in one math class were asked to solve the following problem as their warm-up: $34 + (-3)^3 + (-15) - 32/8$. As students worked, the teacher took attendance and then began circulating around the room. After a few minutes, he quickly asked students to call out their answers in unison. Although there was some consensus, many different answers were shared, and this served as an introduction to the importance of the order of operations when solving math equations.

In English, students entering one classroom were directed to label the next available space in their journals as: *Tell Tale Heart By Edgar Allen Poe: Ten Things I Know About Horror Stories*. Students were instructed to “quick write” ten things that they would find in a horror story and prepare to share their responses after five minutes. It was evident in the observation that the students knew what the expectations were for a quick write, as all of them wrote thoughtfully and with urgency by themselves for the first five minutes of class. The students’ ideas launched the next segment of instruction, where the teacher facilitated students delving deeply into the literary analysis of horror fiction.

In another 6th grade English example, students were directed to go directly to the meeting area and carefully read the prompt that was posted. The teacher said, “Do not begin discussing your thoughts with your partner until you have read the prompt and thought carefully about a specific example from the text.” The prompt directed students

to discuss the following question: *How is the author teaching about man (or woman) vs. society in Maniac Magee?* Student discussions focused on discrimination as a societal issue. In one discussion, a student said, “Maniac was white, but he said he didn’t see the color of people’s skin so he didn’t understand why some people are racist.”

Her partner responded by saying, “When people say they don’t see color, I don’t believe it. I think everyone sees color and a lot of people are racist.”

The other student asked, “But Maniac wasn’t racist, was he?”

“I don’t think so.”

“So, does *man versus society* mean that someone like Maniac is trying to make other people not be racist but it’s hard because most people are racist?”

“I think so.”

After a few quiet minutes of conversation, the teacher led a discussion about the importance of conflict in literature. She provided direct instruction on how to identify conflict as either man against self, man against man, or man against society.

Low Quality: Low Instruction/Low Achievement

In low-quality classrooms, it was very common for students to enter the room talking while the teacher organized the materials and took attendance. Some classes had a warm-up posted or instructions for getting started, but many students did not pay attention to them. It was common in these classrooms for the teacher to provide many reminders about the expectations.

In one of the ineffective math classes, the teacher had given the students an assignment to begin working on as they entered the room. Approximately half of the class was working while the other students participated in various side conversations or

just sat and waited. The teacher was at the front of the room taking attendance. When he was finished, he discussed the study guide that had been given to students and reminded students that their test would include those exact problems. He told them they would have a warm-up quiz on Friday and that the test composed of 15 problems would be on Monday. Next, he said, “Today, we are going to delve into chapter four, and on top of that, we have a benchmark that is coming up as well—Woopie!” The rest of the period was spent going over the problems on the study guide. While the teacher called on one student at a time to give the answers, he wrote them on the worksheet that was projected for the class. Many of the students who had not worked on the problems at the beginning of class simply copied, making it evident that it had not been necessary for the students to do the warm-up. All of the answers were given to students as the class “went over” each problem.

In an English class, the warm-up posted on the board said, “Come in quietly. Sit down. Take out reading log. Take out reading/writing journal. Be ready to learn.” It took several minutes for the class to get settled as students talked and socialized. The teacher became frustrated with the number of reminders that were needed. Finally, she said, “Come on, guys, we do this every day.” There was nothing for the students to think about or do other than organize their materials for the first five minutes of class.

In another English class, students were asked to write in their journals upon entering the room, but no instructions were given about what to write. The level of frustration in this class was evident in the students’ unsolicited comments. At one table, three students were talking rather than working in their journals. One student said, “I wish I could leave this school and work in a taco shop.” The other student responded by

saying, “I am going to write about teachers who can’t teach. It doesn’t matter anyway; he never reads it.”

Scaffolded Academic Dialogue

Scaffolded academic dialogue is sometimes referred to as *partner talk*, but traditional *partner talk* does not always result in teachers providing adequate scaffolding (Kirschner, Sweller, & Clark, 2006; Westgate & Hughes, 1997). When used strategically, scaffolded academic dialogue ensured that all students engaged with the material. Students saw themselves as co-teachers when they used academic dialogue and understood that being able to explain the concepts they were exploring was important evidence that they were learning.

High Quality: High Instruction/High Achievement

In the high-quality classrooms, it was clear from the room environment that students had been taught why academic dialogue was important and what the expectations were for engaging in it. In one math example, the class was discussing absolute value by looking at the problem: $-|-5| + |-34|$. The teacher asked, “When we have absolute value, what are we talking about?” Approximately half of the class responded that it was how far away the number was from zero.

One student asked, “Will it always be positive?”

The teacher responded, “Yes, the absolute value is, but not necessarily the answer. See if your partner can explain why that is.” He gave the class a minute or so to discuss and then said, “Raise your hand if you think your partner had a good explanation.”

He called on a student who said, “Even if you have two positives, if you are subtracting, it might come out negative.”

“So, what is this really asking $-|-5|$?” the teacher asked.

A student responded, “The opposite of the absolute value of -5.”

“Good. This is what I want you to understand. The absolute value will always be positive, but you have to pay attention to the rest of the problem to determine your solution.”

In an English as a Second Language (ESL) class, the teacher consistently had students share their answers with their partners before she began calling on students one at a time. Students understood that they could answer the question first in their home language and that partners could work together to be able to say it effectively in English. In one example, the class had been learning about the setting of a story in both reading and writing. The teacher wanted someone to articulate the meaning of the word *setting*. She asked, “What do we mean when we talk about the setting in a story?”

In one triad, a student asked in Spanish, “What did she say?”

Another student translated the question into Spanish, and the three of them started discussing what a setting was in Spanish. The teacher saw that many of her students were speaking in their home language, so she reminded them that, when they shared out, the discussion would take place in English. A student-made chart displayed in the room illustrated the parts of a story with text in Spanish, Tagalog, and English. The triad got out of their seats and went to study the chart. They coached each other in pronouncing the words correctly, and when the group came back together, they shared publicly with the rest of the class. “The setting of the story is the time and location where the story takes place.” Then the teacher asked the class, “How do you say that in Spanish?” The students translated, and then the teacher repeated the procedure for Tagalog. This teacher did not

speak either Spanish or Tagalog, but she was able to link the content to each of the home languages represented by the students in the room.

Another example of effective academic dialogue came from the lesson discussed earlier having to do with conflict in literature. Students were working in partners to identify the conflict in a story called “Aaron’s Gift.” Two partners had the following conversation: “Okay, so a conflict occurs when Aaron is deciding if he should give the pigeon to his Grandma or to Carl, the gang leader. This is a problem because...” What followed was a discussion in which the partners tried to decide how to phrase their explanation.

One partner said, “He can’t decide which is the right thing to do, so this is man versus self, right?”

“Yeah, I think it is because he is deciding inside of his own head.”

“Okay, you ready? Here’s what we’re going to write. ‘The conflict in the story is man versus self because Aaron is debating within himself what he should do.’”

“Yeah, because he is debating both sides. He’s going back and forth and can’t decide which choice is better.”

“Yeah, it’s kind of like Dr. Jekyll and Mr. Hyde ... like he’s trying to be good, but evil is taking over.”

“You’re right. It is like that!”

Observing the students, the teacher noticed that some needed to do a better job explaining the conflict and how they knew they had identified the correct type. She told them to let her know when they thought they were finished so she could give them some feedback. The partners responded by looking back at their definitions of the different

kinds of conflict. One said, “Maybe we should add the example that no other human being or force of nature is involved.”

The academic dialogue in the classes that were observed from quadrant four was purposeful, on point, and productive. The teacher managed the time needed and adjusted its length based on the responses he or she was observing from the students. In some cases, the academic dialogue was highly structured, whereas in others it served as a quick check of what students already knew or thought about a given concept. A common characteristic was that, in these classes, all students engaged in the dialogue and knew it was a critical component of the learning process.

Low Quality: Low Instruction/Low Achievement

Scaffolded academic dialogue was not present in less effective classrooms. The teacher attempted to engage students in *partner talk*, but seemed unable to structure student discussions in ways that added value to the learning experience. In one class, students were in various stages of writing a story using figurative language while the teacher walked around the room, checking student writing. She asked students to take a moment to discuss with their partners what they had learned from the story they had written. Some students engaged in dialogue, but some did not. Many had conversations, but some of the conversations were unrelated to the lesson. After a few minutes, the teacher said, “If you are finished with your story, you can do your vocabulary.” In this example, the purpose of the dialogue had not been made explicit, it was not scaffolded to include academic language, and there was no follow-up after it occurred to justify having done it.

It was uncommon in most of the ineffective math classes that teachers even used dialogue. In several of these classes, the teacher mainly called on one student at a time to answer questions. It seemed as if the teacher assumed that, if one student could answer correctly, the rest of the class understood the concept. In other words, the teacher failed to check for understanding with everyone and allowed a large proportion of the class to sit passively and copy while a few students participated. In one class, some students were actually yawning as they waited for the class to end. As they continued to go over their math problems, the students were expected to compare the answer they had gotten (if they had done the work) with the correct answer. However, the teacher did not ensure that students were participating, so the disengaged students were left with incorrect answers remaining on their papers. Using academic dialogue as a tool for engaging students in being analytical about their answers would have increased the level of engagement and the level of learning for a majority of students.

Real-world Connections

Real-world connections were prominent in all of the classes that exhibited high instruction and high achievement. Sometimes they served as a springboard for studying a concept in depth, while at other times real-world connections were woven into the tasks teachers assigned. The important thing was that students connected the concepts they were learning to real-life experiences that made sense to them.

High Quality: High Instruction/High Achievement

In math, an example of the effective use of real-world connections was observed in the teacher's introduction. The teacher connected the topic for the day directly to the lesson from the previous day. He said, "Yesterday, we learned about discounts. This is

important because we all like going to stores and seeing things on sale. Today we are going to learn about the opposite: how businesses make money.” Then, the teacher created a scenario that the students understood. He held up the remote to the Promethean board and said, “This is a universal remote called the magic wand that your company produces at a cost of \$50. Your business wants to sell all of these magic wands to the school principal because all of the classrooms have Promethean boards that they work perfectly with. As a business owner, how much should you sell them for?” The groups immediately began discussing the price, and as the lesson progressed, the teacher continued to build in new learning.

In an English example, the teacher provided meaningful background information to students. In the lesson discussed earlier about horror stories, the teacher said the following to the class after they shared their quick writes: “For the next couple of minutes, I just want you to listen. While I am talking, I want you to note three important facts that you learn about the author.” He then said the following:

Edgar Allen Poe was born in the early 1800s. He lived to be in his forties and had a sad life. He didn’t get along with his father. They fought all the time. He battled with depression and felt bad his entire life. He died penniless, battled with alcoholism, and nobody ever read any of his writing. I wrote the word *posthumous* on the board, which means that he didn’t get any recognition for any of his writing until after he died. He was really the first person to write scary, suspenseful stories. If you were the first person to write these kinds of stories, what do you think people would think of you? He was also a poet and wrote a famous poem called

“The Raven,” which is what the Baltimore Ravens football team is named after.

When he finished, he added, “Now write three things about his life that you remember.” The students immediately began writing in their journals, and then the teacher introduced the story, “The Tell-Tale Heart.” In a very short amount of time, the teacher made the author into a real person that students could know and connect with.

Another example of real-world connections comes from the same lesson. The teacher introduced the class to the Anticipatory Guide and explained that the purpose was for students to see how their own life experiences influence both how they comprehend stories and also the writing they create. The Anticipatory Guide included five statements that students needed to read and then determine whether they agreed or disagreed. The teacher said, “I need you to think about each of the points and then write your opinions about it. You need to either agree or disagree and provide a rationale that explains your thinking.” Some students worked alone, while others worked with their neighbors on the following statements:

1. *People who are insane don't always know that they are insane.*
2. *Sane people sometimes imagine that they hear things.*
3. *If you commit a major crime, sooner or later you will be caught.*
4. *When you've done something wrong, it is agony to wait to see if you'll be caught.*
5. *All people share the same fears.*

As students decided whether or not they were in agreement with each of the statements, they were compelled to ask each other for clarification if needed.

“What exactly does it mean to be insane?” one student asked her neighbor.

“I think it means you’re crazy.”

“I thought so, but I just wanted to make sure. So for number 2, it’s talking about normal people who hear voices, right?”

“Yeah, I think so.”

“What did you put for that one?”

“I put....” She looked down at what she had written and read, “I disagree because only insane people hear things that aren’t really there.”

“Okay, I was thinking since it said ‘imagine’ that I would say agree because lots of people can imagine things even if they aren’t really happening.”

“What do you mean?”

“Like I could imagine you said something to me, but you really didn’t say it. I could just think about if you said it.”

“Yeah, that’s true. Like if you’re imagining how something is going to go so you can plan for it or something?”

“Yeah, I think normal people do that a lot.”

Here, students were interacting with situations that required them to link to a personal situation as they decided whether or not each statement might be true. The teacher was preparing students for issues that would come up later in the story, with the knowledge that if students were invested in the issue, their comprehension would be enhanced.

Low Quality: Low Instruction/Low Achievement

In math, the least effective teachers rarely gave students opportunities to connect the numbers, quantities, and operations they were using to anything in the real world. In some instances, the class would spend the whole period doing “naked” computations where students had to manipulate numbers and calculate solutions without any connections to the real world. In one math class, the students confronted a word problem that involved buying a car that cost \$10,000 plus 6.1% sales tax. The teacher called on students one at a time to work through the steps, but at no time were the quantities connected to anything in the students’ lived experiences. In fact, the teacher skipped over the words and went directly to the computation.

“For problem 1, if we have 6.1%, how many places do we push the decimal?”

The teacher called on a student, who said, “Two?”

“Right. Now do we push it to the right or the left?”

“To the right?”

“Actually, that will make the number bigger. We want to push it to the left. If we push it to the left, what will we have?”

The teacher modeled while the class watched and some students called out, “.061.”

The teacher acknowledged the students’ answers and then continued, “Okay, now we can multiply. Now, what is $10,000 \times 0.061$? How many places, exactly, do we push the decimal if we are multiplying by 10,000?” The students did not provide any answers, so the teacher said, “Four times, right? So that’s 610 dollars. Now how much is the total price of the car? \$10, 610, right? Do you all understand how we did that problem?” He

then retold the procedure to the class, with no connections to the process of actually buying a car at a base price plus tax.

One less effective English classroom had two teachers. One of them was at the Promethean board while the other one walked around the room checking on students. The teacher at the front of the room said, “We are going to do a quick write. Answer this question in a short paragraph with complete sentences. What do you know about coral reefs? You might include your thoughts on why they are important, where you can find them, and/or what is happening to them.” Then, the other teacher added, “You need to write in complete sentences, without talking!” She kept repeating the words *coral reef* and spelling them for students. As some students worked and others engaged in other various off-task activities, the teacher called out, “Three more minutes.” All students began writing, and as some students finished, they sat and waited. The teacher who was walking around said, “If you don’t have ideas, try to get at least something down and put it in a good simple sentence.” The students never had a chance to share their writing or discuss their own understanding of and experiences with coral reefs. The teacher at the front of the room launched immediately into a series of pictures he had collected from a scuba trip he had taken. He described the coral reef and then said, “Today we are going to learn about expository text.” The other teacher began passing out a page of text entitled “Magical Coral Reefs” and asked the students to identify the text features. Although the teachers attempted to create a context for learning about expository text by introducing students to coral reefs, the students did not develop a connection with the text because of the way it was presented.

Front Loading

Front loading is a strategy often used with English Learners who struggle with academic vocabulary. The teacher “front loads” the vocabulary by providing meaningful experiences with the words before students are confronted with them in the text, but front loading can be helpful in other instances as well. Front loading allows the teacher to set clear expectations for students and prevents the pitfalls that students often experience when they encounter challenging or complex tasks.

High Quality: High Instruction/High Achievement

High-performing teachers always described exactly what they expected students to do. They did this in a way that made the task richer and more meaningful while at the same time scaffolding the task so that students felt supported in completing it successfully.

This first example came from the math class where students were learning about discounts. The teacher had assigned groups to come up with what they thought was a reasonable price for selling a remote control. As the teacher noticed that all of the groups had made the decision about their price, he said, “When I call on your table, I want to know two things: How much should we charge and what’s your rationale?” The teacher modeled the language for reporting out by saying, “Here is how it will sound. You are going to say, ‘We think ___ is the best price. Our rationale is....’”. He then gave the students time to fit their ideas to the framework he provided and then called on each group to share their responses.

In the lesson about Edgar Allen Poe, the teacher used front loading in two separate tasks. He said, “What we are going to be confronted with in this story is that we

don't really know if we can trust the narrator. By the end of the story, I want you to judge him and determine whether or not he is a good person." The students each received a paper entitled "Exploring Sensory Imagery with Edgar Allen Poe's Tell-Tale Heart." The teacher said, "As we read, I want you to think about what the author does to suck us into the story, so that when you write your story, you will be able to use sensory imagery to keep your readers involved. Which sense do you think is going to be the most important?" The students provided many suggestions, and then the teacher asked the class to open their textbooks to page 523. This form of front loading gave students a specific purpose for listening as demonstrated by the notes many students made of instances where sensory images dominated the writing while the teacher read the story. When the teacher stopped the class and had them visualize what was happening, he gave students opportunities to actually feel the effects of powerful writing.

Low Quality: Low Instruction/Low Achievement

Although front loading is an effective strategy that is artfully used by the very best teachers, there was no evidence of front loading in any of the low-quality classes in either math or English. In one math class, the teacher was showing the students how to work through a very complex problem. He was frustrated because the students did not understand. He drew three graphs on the board and asked, "What do the lines show us?"

One student answered, "Solutions?"

"Right, they tell us all of the solutions to the problem. But, on the third one, the points did not all fall in what we call a linear pattern, right? Why is that?"

The teacher called on a student who said, "The x has a little two."

“Right, it was a different kind of function, so it wasn’t a straight line. It grew in a different way. Remember, it went, 0, 1, 4, 9, 16. What is the difference between the growth rate in a linear versus non-linear function?” The teacher displayed two different data sets and asked, “Do you see that this one does not have a constant rate of growth?” The students did not see what the teacher was talking about, and the teacher struggled to break the concept down in a way that linked their current understanding to the advanced understanding he expected. The teacher used very advanced vocabulary and also presented a complex problem that the students were not yet able to access. Rather than adapt the problem, the teacher resorted to doing the work for the students and assumed that once they copied the procedure, it was okay to move on to the next concept.

In one English class, where the teacher was instructing the students about how to understand character traits, the teacher said,

We learn about character traits from the characters actions—what they do, what the character says, and what other characters say to and about them.

Does that make sense? Don’t write it down if it doesn’t make sense. Does it make sense? That’s a tough one—what other people say about them.

Students copied the three methods for understanding character traits into their notebooks, but never got an opportunity to apply the skill to a piece of text. If front loading had occurred, the teacher might have provided a text and before reading, she might have said, “While reading this passage, I want you to pay attention to what you learn about the character. When you think you have something, think: how did I know that?” After reading the passage, the teacher would elicit strategies from the class and then present the

required methods if they hadn't already been shared. As students shared specific examples, the class would identify the method that was used.

Differentiated Instruction

Differentiated instruction implies that students at different ability levels or with different prior experiences will engage with the concept differently. Knowing the strengths and needs of students enables teachers to plan activities that all students can access. Differentiated instruction is also a form of scaffolding. Scaffolding is the process in which new information is presented in a manner that allows the learner to connect and integrate prior knowledge with the concepts that are still unfamiliar. In order to provide effective scaffolding, it is important that the teacher understand both the preconceptions and misconceptions that different children bring to the learning process and adjust the task accordingly to meet the needs of students at a variety of levels.

High Quality: High Instruction/High Achievement

All of the high-performing teachers structured their classes to allow for individualized and/or small group interactions throughout the lesson. In a math example, the teacher was trying to help students understand what it meant to isolate a variable when solving equations. First, he told a story about someone in the class feeling isolated and then led a discussion on the meaning of the word. Then, he put a series of problems on the board that ranged in difficulty from easy to more challenging. He asked students to work through the problems either by themselves or with a partner. They were to make up a story to match the equations and then match the numerical representations with the words in their story. First, he walked around the room to make sure that everyone knew what to do. His more proficient students breezed through the first few problems, and as

the numbers became more complicated (using decimals and fractions as opposed to whole numbers), the students were able to apply the process they had gone through with the easy numbers to create more difficult scenarios. The teacher made sure struggling students had access to a more proficient peer and reminded them that they should check their ideas with each other if they were not sure they had it correct. Then, he sat down with one student who was solving the equation: $m - 5 = 3$. He said, "Tell me your story."

"I had a bag of candy. I gave five away, and I was left with three."

The teacher asked the student's partner, "What is her question?"

The student responded, "How many candies did she start with?"

"To figure that out, what do you have to do?"

"Find out how many she started with?"

"Okay," coached the teacher, "how can you show that using numbers and symbols?"

The student said, "I am going to pretend that I am giving them back right here..." On the paper, the student added five to the three. "And then if I gave back the five, that means I never took anything away from the m ." The student added five to the side of the equation with $m - 5$. "The whole equation was $m - 5 (+5) = 3 (+5)$."

"Good. So, why do we call that 'isolating the variable'?"

The students studied their equation for a minute and then one of them said, "Because now the m is by itself because the $-5 (+5) = 0$, so $m = 8$!"

"Good, and what does the eight stand for?"

"The original bag of candies."

"Great, now try the next one."

By providing the whole class with a range of problems, each student was able to settle in on a problem that was at their “just right” level of difficulty. The open-ended nature of the task allowed the teacher to check in with each student or pair of students, and he was able to provide differentiated support right at each student’s point of need.

In an English example, the teacher asked students who were finished to come to the meeting area. He told the rest of the class they had 10 more minutes of writing time. Eight students met him in the meeting area. He told them that if they were completely finished, all they needed to do was to show it to an adult and have the adult read it. He excused them to work on comic book versions of their story while the rest of the class continued to work. In the comic book versions of their stories, students were taking what they learned about plot planning and story structure and applying those skills in a new genre. While the most proficient students were sufficiently challenged and motivated by this next step in the writing process, it freed the teacher up to continue to provide support for students working at other levels in the process. In a subsequent conference, he scaffolded the editing process for a pair of students by giving them explicit directions for how to critically read each other’s work and then the teacher worked individually with an English Learner who was analyzing his text for consistent tenses.

Low Quality: Low Instruction/Low Achievement

Differentiated instruction was not observed in any of the low-quality classrooms. In these classrooms, the instruction remained on the whole-class level. If students did work independently, either the teacher told students how to do the problem if they had difficulty or the students simply copied from a more proficient peer. The tasks these teachers assigned to students were not conducive to observing student thinking, but

instead focused on computation and getting the correct answer. If the student had the correct answer on his or her paper, the teacher assumed the student understood the concept. Because all students were given the same task and completing the task failed to reveal the students' thinking processes, differentiated instruction could not occur, and the students were not able to benefit from the process of instructional scaffolding.

Feedback/Conferring

Another common characteristic in the classes that exhibited both high instruction and high achievement was that the teacher was actively involved in providing immediate feedback and coaching to students as they tried on the skills they were learning.

High Quality: High Instruction/High Achievement

High-quality teachers made a point of constructing tasks that allowed them to observe children's thinking. There was a sense that all of the students knew that the teacher cared about what they thought and that any misconceptions would be recognized and addressed. All of the effective math teachers that were observed used whiteboards or individual personal computers during whole-group instruction. For example, in one lesson about integers, the teacher said, "Show me on your whiteboards what you got for this problem: $-23 + (-15)$." Students showed their answers: -38 . The next part of the expression was " -32 ," and the teacher asked, "What do I take away?"

Many students in the class responded, "32 positives!"

The teacher asked, "Can you?" Most students said no, and the teacher asked, "Is it because you don't have any positives or because you don't have enough?" Some students said, "Not enough," and the teacher sensed that the whole class was not following along. He said, "Show me on your whiteboards what you are going to do." As

students got their answers, they held up their solutions, and the teacher either gave them a silent thumbs-up or went over to them to provide guidance or asked them to check with someone else. If they got a thumbs-up, they either went on to the next problem or helped a partner who was having difficulty.

One student who had thought the answer was -6 changed his answer to -70 by subtracting 32. The teacher said, "Do you understand your mistake?"

"Yes."

"Tell me."

The student said, "I took away the 32 from the 38 as if the 38 was positive."

The teacher said, "Right. When you take the test, they are going to try to trick you like that, so what do you have to do?"

The student thought for a minute and said, "I have to pay attention to the signs."

The important aspect of feedback was that students always seemed to know they were going to get it and even expected it. They also knew that it was critical that they follow the instructions carefully so as not to disappoint the teacher. In these classrooms, there seemed to be some underlying agreement between the teacher and students that their work was worthy of both the students' and the teacher's attention.

Feedback was also a critical feature of effective English classes. In one of the lessons mentioned earlier, as students worked independently, the teacher viewed student writing on his computer screen. Based on an initial scan of all students' writing, the teacher decided to confer with a student about his paragraphs. The teacher provided very specific instruction about how to analyze each paragraph for its main idea. Next, the teacher met with two students who were learning how to help edit each other's work.

One was reading her story to the other as the other one checked for accuracy. The first student read, “The Day of the Bread (Title). Now that I look back at what happened, I think that was the most interesting day I ever had at school. I will never forget Mr. Fran.” Each of her sentences was color-coded to match the expectations for how the story was to be structured. The teacher had created a checklist the students had followed to make sure they included the essential elements of a fiction story. The teacher asked the listener, “What kind of comments did you have regarding the first page?”

“It was mainly punctuation.”

“See how he has a lot of dialogue?” the teacher asked. “Now look at yours. You begin yours with a lot of narration. You can help him with his narration, and he can help you add more dialogue. For example, where does his story take place?”

“Arizona.”

“Yes, but he says nothing about how hot it is or what it looks like. I have no idea from his writing about what any of the scenery is. You can help him with this.” As the teacher left, the students started helping each other with specific revisions based on each of their strengths and needs.

Low Quality: Low Instruction/Low Achievement

In the low-quality classrooms, feedback was either in the form of publicly addressing student misconceptions with the whole class or responding to students in ways that did not generate deeper understanding of the concepts they were learning.

In math, the teacher asked the class how to change $\frac{4}{5}$ into a decimal. “What do we have to do in order to convert a fraction to a decimal?”

One student said, “Divide?”

“What do we divide?” The class remained silent. “Are you taking notes and writing these things down?” The teacher called on a volunteer student, who said, “You should multiply.” The teacher said, “No. What we do is divide 5 into 4. But 5 won’t go into 4, right? So, what do we have to do?”

One student said, “Add a zero?”

“Right. See. How many times does 5 go into 4? Can’t do it, right? So, how many times does 5 go into 40?”

The same student said that the answer was eight, and the teacher said, “Exactly right. The decimal is going to be 0.8. I need you guys to understand this.”

In another class, students were working on an assignment and submitting their answers via Active Engage, which is a computer program that provides immediate feedback by telling students if their answers are correct or not. The teacher monitored the class and assisted students who were not completing the procedures correctly. Students were required to show the steps as they worked through the problems. One student figured out mentally that $3 - 4.7$ would be -1.7 . The teacher said, “I need you to show me the steps.” The student tried to complete the problem the way the teacher had modeled it, by subtracting the smaller number from the bigger number, but instead the student subtracted 4.7 from 3.0. The student made mistakes on the regrouping and got 2.3. Then, the teacher went to the board and reminded the whole class that they had to start with the bigger number and then look at the sign to figure out whether the answer was positive or negative. In this example, even though the student had used his number sense to obtain the correct answer, the teacher was unable to connect his original thinking to the

procedures she was teaching or help him understand why his computation did not match his original answer.

In one 8th grade English class, the students were each assigned one vocabulary word to study and then present to the class. The students had to present the part of speech, the definition, a sentence in which the word made sense, and an illustration of the word. While students shared, the class was directed to write down the definitions on their own papers. One student shared his work on the word *stern*. He stated that it was a noun and read his sentence aloud: “The teacher sternly told the class to be quiet.” His picture showed a teacher yelling at the class with an angry face and a pointing finger. The teacher did not notice that he used the word in its adverb form. Another student said, “I thought it was an adjective.”

The teacher said, “It’s okay to put both so you know that words can be used in different ways.” The class never explored the meaning of the word as a noun or how the word was changed to an adverb. The original student did not revise his work; thus, the feedback the teacher provided did not facilitate deeper understanding.

Structured Reflection

Reflection proved to be a critical component of the learning process. High-quality teachers required that students think on their own and reflect on their work in order to improve their understanding. Reflection was not something that happened after the learning; rather, it was an integral part of the entire process.

High Quality: High Instruction/High Achievement

In one math class, the teacher wanted students to reflect on the meaning of integers. He asked the class to look through their notes over the last five days and think

using the following questions: *What are integers? What have you learned about adding integers, multiplying integers, dividing integers, and subtracting integers?* He gave students five minutes to study their journals. Many students also referred to the word wall as they crafted their responses to the teacher's questions. One student whispered to his partner, "Can we take a negative number away from a positive? What do we need to do?"

She told him, "You have to create a neutral field. It is in the notes from Thursday."

After five minutes of "studying," the students turned to their partners and summarized what they had learned about integers in relation to each operation while the teacher circulated and listened in on student conversations.

In the lesson about the remote controls, after the groups had shared their ideas, the teacher said, "I want each of you to think about this by yourself before you start talking to your group. Every business wants to make money, so if it costs me \$50, which businesses here..." The teacher paused to gesture to a chart he had created with the answers from each of the nine groups. "...will possibly end up losing more money rather than gaining and why? Once you have your list, you can begin discussing it with your group." This structured reflection required all students in the class to generate ideas on their own before sharing answers aloud. It also added to the level of critical thinking that the group engaged in because they compared each other's ideas, critiqued each other's work, and synthesized all of the answers into one list.

In many of the classrooms where writing was the focus, students were finishing, reflecting on, or extending their writing. In one class, the teacher created a reflection sheet that asked the following questions:

1. *What is your hook? Write two sentences to describe it.*
2. *When does your story take place? Write the sentence that describes it and place #1 by it in your draft.*
3. *Describe the plot and show on a timeline what the rising action is.*
4. *What is your climax?*
5. *How does your character feel? Write the sentence that describes how they feel at the climax. Put #2 by it.*

One student was reading over his final draft and making additional changes as he responded to each question. He asked a person at his table, “Do you think this is true?” He read what he had written on his reflection sheet: “The character’s flaw leads to the climax because of all of the fighting.”

The partner responded, “What caused the fighting in the first place?”

He explained what happened in his story, and they worked together to find the place in the story where he could change the dialogue so that the character’s feelings were visible.

In this 6th grade class, reflection was used to foster students’ critical analysis of their own work and to link the stages of the writing process with specific content objectives. Students applied the literary techniques they were learning about in reading to their own writing and then connected back to the content of the lesson. Students used the academic language of the discipline in a natural way to interact with literature, each other, and their own writing.

Low Quality: Low Instruction/Low Achievement

Structured reflection was not evident in any of the low-quality classrooms that were observed, but some teachers used the word *reflection* in an attempt to have students think about their work. In one math class, the teacher opened with the following lecture as students worked on the seven problems that had been assigned as their warm-up:

We have three things to do today. In a minute, we're going to play a game called *What Did He Say Yesterday?* I am going to start asking you what the key point of our learning was the day before and have you reflect that back to me. Each group will be responsible for choosing one representative to share with the class. After that, we are going to go over the warm-up problems, and then we are going to start reviewing for the benchmark assessment.

After the introduction, the teacher went to his desk and began taking attendance while some of the students worked on their warm-up problems and others either chatted or slept. When he finished taking attendance, he began calling students to the board one at a time to show the class how they solved each problem. A total of eight students participated in modeling the warm-up problems for the class while the rest of the students were passively listening, sleeping, chatting off-task, or playing. After the warm-up problems were complete, the teacher passed out an unrelated worksheet and then noticed the class was over. He ended by telling the class they would review for the benchmark the next day.

Lesson Closure and Connected Homework

Lesson closure means that the teacher brings the class together as a group and facilitates a synthesis of the big ideas of the lesson. It is a result of the teacher paying careful attention to the different ways students interacted with the concepts of the lesson so that the best thinking could be made public and a bridge could be built between yesterday's, today's, and tomorrow's work. Connected homework means that assignments were directly related to what students worked on in class that day and served as an extension or reinforcement of what was learned.

High Quality: High Instruction/High Achievement

A common characteristic of the lessons observed in high-quality classes was that the teacher took time at the end of the lesson to emphasize the learning goals and tell students how the work they did in class was related to the work they would do for homework. In most cases, students' homework served as their "ticket into class" the next day. This made the teacher immediately aware of students who had not completed it and enabled him or her to address those students directly. Almost always, students were required to get their homework signed or explain their work to their families.

In math, at the end of a lesson about graphing linear inequalities, the teacher said, "There are three parts to your homework tonight. We explored graphing linear inequalities in class today, but now I want you to read the textbook to see how it describes the process. The textbook provides instructions for diagramming and gives a real-world example that I think you will relate to. Read the text carefully, take the quick quiz, then try to explain to your family what you know about graphing linear inequalities. They probably

are not going to know what the words even mean, so you are going to have to explain it. We'll discuss it more in depth tomorrow, so I want everyone to complete all three parts: Read, Quiz, Explain.

In the ESL class mentioned earlier, the teacher closed the lesson by bringing the class together as a group and asking them to reflect on their writing. She displayed the following sentence stems on the board and asked students to work with partners to answer the following questions:

1. *What is the setting? The setting of my story is....*
2. *What is the plot? My story is about....*
3. *Who are the characters? There are ____ characters. They are....*
4. *What is the point of view? The story is told from the perspective of....*

The expectation was that all students use complete sentences. After the students had reviewed their work with a partner, the teacher said, "Circle five words—one from each section—that you need to learn in English to improve your writing." She waited as the students circled their words and then said, "Look them up tonight so you are confident in their use. Teach them to a family member and talk about why you are using them at school. Tomorrow, you will spend the first five minutes of class integrating them correctly into your story."

In one of the writing classes, the teacher ended the lesson by bringing all of the students to the meeting area and saying,

Tonight when you are looking over your final draft, I want you to take your paper and put it in front of your mom, dad, grandma, grandpa, any

What stood out in these classes was that each teacher took time at the end of the lesson to emphasize the important concepts he or she had taught so that the big idea could be solidified in each of the students' minds. The homework was directly related to the ongoing work students were doing in class and whether the students completed it or not would have a direct impact on the work that was planned for the next day. In many cases, the homework was individualized so it related specifically to what each student needed to facilitate their own progress. Finally, it facilitated communication and collaboration with family members, creating a natural and ongoing connection between home and school.

Low Quality: Low Instruction/Low Achievement

Lesson closure was rarely observed in the low-quality classrooms. Often, the class would finish the work for the day and pack up and wait for the bell to ring. In some instances, the class would work until the bell and the students would just get up and walk out as the teacher called out various directions to the class like, "Remember to study for your test!" Or, "Sit down and wait for me to excuse you!" Or, "Good work, see you tomorrow!" Although homework was assigned in most classes, the degree to which it was integrated with the work for the day varied on a wide range. In some cases, the entire lesson the next day was centered on going over the answers to homework problems, especially in math. In other cases, there were packets assigned that were given out on Mondays and due on Fridays. In still other cases, the warm-up for the day was to copy the homework assignment into student planners. These methods provided evidence that the teacher had planned what to teach each day and had determined in advance the work that would be completed at home. However, generally speaking, in low-quality

classrooms, homework was not strategically or explicitly connected with the instruction that occurred in class.

Integrated Strategies

Often, the teachers in high-performing classrooms would use a combination of strategies. For example, in the math lesson about integers, the teacher gave the students the following problem to solve: $56 - 78 - (-32)$. He asked the class, “What sign is our answer going to be?” The class answered that the answer was going to be negative, and then the teacher said, “Tell your partner how you know it is going to be negative and give a real-world example to prove it.” Some students began solving the problem, and others began working together on their real-world examples. The teacher provided feedback in the form of either affirmations, questions, or direct coaching about how to think through the problem. One student was having a hard time figuring out the answer. The teacher asked him a series of questions, which guided him through the process.

Teacher (T) - “What do we do first?”

Student (S) - “ $56 - 78$.”

T: “What do we have?”

S: “56.”

T: “What are we taking away?”

S: “78.”

T: “Is it that you don’t have enough or that you don’t have any?”

S: “Don’t have enough.”

T: “Use your white board to show me how many you want.”

The student subtracted 56 from 78 and got 22.

T: "So, what are we going to do?"

S: "Put 22 neutral fields." The student wrote (+22 & -22) on his paper."

T: "Good. What is the problem asking you to do?"

S: "Take away the +22 and take away the +56."

T: "Good. You took away 78. What do we have left?"

S: "-22."

T: "Now, for $-22 - (-32)$ what do we need?"

The student said to the teacher, "Okay. I got it. I can do the rest."

The teacher went to listen in on a triad of students who were trying to figure out what each number could mean for their story. They decided to use money and had started with 56 dollars. They decided to buy an iPod off of e-Bay for \$78 and now had \$-22, which they had borrowed from their mom. They were arguing about what it meant to take away -32, when the teacher asked, "What might have happened that would require you to take away a negative?" One student said, "If you take away a negative, it means the negative isn't there anymore, so maybe the kid got some birthday money or something." The teacher responded, "Good. How much did he get?" The group argued some more about why someone would get \$32 as a birthday gift, but decided to use it for their story anyway and ended up with a final answer of \$10. In this example, the teacher used scaffolded academic dialogue, real-world connections, and feedback all within a five-minute lesson segment.

Ineffective Practices

There were several common practices in low-quality classrooms that either inhibited active engagement or resulted in students' active disengagement from the

learning process. These included: (a) students waiting with no academic expectations; (b) students copying from the board, texts, or each other; (c) public reprimands for off-task behavior; (d) calling on students one at a time; (e) rapid-fire questioning; and (f) teachers answering their own questions. None of these characteristics were observed in the high-quality classrooms.

Students Waiting–No Expectations

Unstructured time was problematic in many of the low-quality classrooms. It occurred primarily at the beginning and end of classes or whenever a teacher gave the whole class the same amount of time to complete a task with no extensions for students who finished before the time boundary. In one example, the teacher was providing extra time for students to finish a test that was given the day before. At the beginning of class, the four students who had completed the test were told to wait until everyone else finished. They were not given an assignment or any direction about how they should spend their time. As other students finished the test, they turned it in to the teacher, who sat at his desk at the front of the room on his computer. By the end of the period, two students were still working on the test while the 27 others decided on their own to engage in a variety of activities. Some read or completed assignments from other classes. Some put their heads down or had quiet side conversations. Some ate candy, shared snacks, or wrote notes to one another. Two students continually flicked items at classmates, causing them to yell out. Only then would the teacher look up from his computer and say, “Come on, guys, no talking.”

Other ineffective classroom practices that resulted in students having to wait occurred when the teacher needed to take care of classroom business like taking

attendance, checking homework, or passing out materials to one student at a time. All of these practices promoted off-task behaviors, increased the potential for severe discipline, and even jeopardized student safety.

Copying

Copying answers, definitions, procedures, or notes was often perceived by ineffective teachers to mean that students understood the concept. There were many examples in the low-quality classes of students turning in work that had been copied in class. Often the teacher would call on students to give answers, the teacher would write it down if it was correct or correct the student if the answer was wrong, and then the rest of the class would copy what the teacher wrote. Some teachers spent the entire period having students copy and even complimented the students at the end of the class when the “work” was collected.

Public Reprimands for Off-Task Behavior

When tasks were not structured for student success, resulting in off-task behavior, the classroom environment was severely affected. Throughout the lesson about coral reefs, the teacher would publicly admonish students. In one lesson segment, the teacher said, “Now I am going to show you some slides of what a coral reef looks like. Not talking out loud, but looking at the picture, I want you all to come up with your own opinion. What do you see in the picture? (Student name), you need to listen! What do you see that is kind of unusual? Not out loud! In your head, to yourself!” The teacher continued to show slides, occasionally calling out admonitions such as, “(Student names!) We’re not talking to one another. Turn down the volume and just look at the pictures!”

After seeing some examples of coral reefs, the teacher said, “Now you need to write down a prediction about the article and then sit quietly and read it by yourself. There should not be any talking. That means people like (Student name) are not talking. (Student name), you need to be reading too.” While the class read, the teacher continued to repeat, “Everyone is reading. Sit quietly and read. Everybody is reading the whole thing.” The teacher walked around the room talking loudly to students, making it very difficult for students who were trying to read to concentrate. The teacher even said, “(IEP Student’s name), I am going to give you a different article.” Some other students were continually off-task and disruptive, but the teacher never intervened at all which resulted in many students receiving mixed messages about the teacher’s expectations.

After the students had completed their reading, the teacher said,
Now, you need to compare your prediction to what is actually in the essay.
Really look at your prediction and tell me if you did a good job making a prediction. (Student name), you better not be writing notes! Give that to me. If you are writing notes in class, I’m going to start calling parents.
You have one more minute to write. Okay! Everyone needs to be seated.
You do not need to be coming up and giving me anything. Shh! Just wait a second! Hey! Okay! (Student name), where’s your paper? Everybody sit down. Sit down! We’re not going anywhere! Okay! (Student name), tell me one thing you learned.

The student said, “They’re dying.”

The teacher, distracted by the behavior of another student, called out, “(Student name)! I am going to call your mother because you left your seat without permission!”

This level of negativity by the teacher was not uncommon in low-quality classrooms. Rather than focusing on content, the majority of the teacher’s energy was dedicated to ineffective management of student behavior.

Calling on Students One at a Time

This ineffective practice is the antithesis of scaffolded academic dialogue and resulted in the majority of students disengaging from their learning tasks. In many of the low-quality classes, the teacher would ask a question and, although at first many students would raise their hands, the teacher would call on one person and then have an extended discussion with that one person or simply go on to the next question and call on another person. When the teacher called on only those students with their hands up, then students without an answer were essentially denied the opportunity to think and make sense of the material. The teacher excused them from thinking and failed to provide the supports needed to engage them with the material. There were many instances where the one-to-one dialogue between the teacher and one student at a time extended through the whole period, resulting in many students sitting quietly without ever saying or doing anything.

Rapid-Fire Questioning

Another more stressful version of calling on one student at a time occurs when a teacher prefaces a command with a random student’s name. In some classes, the entire period was spent with the teacher going through a series of questions like the following:

Teacher (T): “(Student name), give me all of the factors of four.”

Student (S): “1, 2 4.”

T: "What about the factors of 8?"

S: "1, 2, 4, 8."

T: "What about 12?"

S: "1, 2, 4..."

T: "No."

S: "...3, 4, 6, 12."

T: "Good. So what is their greatest common factor?"

S: "4."

T: "Right. Correct. Very good. (Another Student), what process are we talking about when we have $0.2 + 0.2$?"

Student 2 (S2): "Adding."

T: "And what is $0.2 + 0.2$?"

S2: "0.4."

T: "Exactly. And if we have 0.2 times 0.2, what process is that?"

S2: "Multiplication."

T: "And what is 0.2 times 0.2? How many places are we going to move the decimal?"

S2: "One?"

T: "No. $0.2 \times 0.2 = 0.04$. So, what do we have to do if we are comparing?"

S2: "Subtraction?"

T: "And what do we get if we subtract 0.04 from 0.40?"

S2: "0.36?"

T: "Exactly right."

In this class, the entire period was spent in this form of rapid-fire question mode as the teacher went over problems that had been assigned the night before. The teacher would write the answers on the worksheet, which was projected so the students could watch, and students were expected to check their answers.

Teachers Answering Their Own Questions

When assessing critical thinking, the following examples show why it is important to look not only at the questions that teachers ask, but the responses from the students followed by the response from the teacher addressing the students' answers or failure to answer. In almost all of the low-quality classrooms, when the teachers asked questions that the students could not answer, the teacher would answer the question for the students.

In one math class, the teacher asked several questions without receiving any student responses. The teacher said, "Let's do #4. It says, 'What's the average of $3\frac{3}{10}$, $3\frac{4}{2}$, and $1\frac{1}{5}$?' What's another word for average?" The students did not answer, so the teacher prompted, "Starts with an *m*." When the students still did not respond, the teacher said, "It's mean, right? Mean is another word for average. So, how are we going to figure this out?" After a period of silence, the teacher continued, "We have to add them up, right? So, what do we do to add up the fractions?" The students sat impassively until the teacher said, "We have to get a common denominator, right? What will that be?"

Finally, one student volunteered, "10?" The teacher answered, "Right," and then showed the students how to add up the numbers to obtain a total of 12 ($3+3+4+1+10/10$). "Now, since we are finding the mean, we divide by 4." The teacher wrote $12/4 = 3$ on the

board. When he asked the class what the answer was, a few students orally read the answer he had given.

An English example occurred as the teacher was calling on students one at a time to read. The teacher would stop at certain points in the story and ask questions. At one point, he stopped the reader and asked, “What does *interchangeable* mean?” Many students had not been following along, but of the ones that were, no one offered an answer to the teacher’s question. Instead of engaging in a discussion, the teacher said, “It means the parts can be interchanged or switched with each other. You all should know that by now. We discussed it yesterday.”

When the teachers in these low-quality classes answered their own questions for students, they often assumed afterwards that the students understood and moved on. In the English example, the teacher was checking for understanding and found out that either the class was not paying attention, or they did not understand the concept. It was clear from his response that he had “taught” the concept before, but the students had not acquired the knowledge he had intended. When their misconceptions became evident again, he repeated the same ineffective strategy of teaching by telling and never addressed the underlying issues that lead to the students’ misconceptions in the first place.

When teachers answer their own questions and assume students understand, they fail to engage the class in active learning. For example, one teacher asked the class, “How did we get 19 for an answer here?” He called on a student, who began to explain his thinking, but the student paused to think about what he had done. Before letting the student think, the teacher interrupted and said, “You figured out that it was four rods for each length and then subtracted one, didn’t you?” There were no students who could

explain how the answer of 19 was derived, and after the teacher told them, he continued to teach as if the class now understood.

Distribution of Quality by School

The graphs below show the distribution of quality instruction in each school. As previously mentioned, quality for the purposes of this study is defined both by the teacher's instructional score and the achievement of grade-level standards by students. School #1, which has been struggling to move out of program improvement status, had only five classrooms that achieved high instructional scores and high achievement compared to 20 classrooms in School #2. At the same time, high instruction and low achievement (quadrant 3) was more likely in school number 1. This may have been attributed to teachers teaching well, but not keeping up with the curriculum pacing guides, which resulted in many concepts remaining untaught. Furthermore, 21 classrooms in School #1 fell into quadrant one representing low instruction and low achievement compared to only 13 classrooms in School #2. Additionally, 13 classrooms at School #1 achieved improved test scores despite having poor instructional scores (quadrant 2) compared to only 7 in School #2. These results align with the comparative Academic Performance Index (API) for each school. The API is the school's state-wide ranking based on student academic performance and progress. The API ranges from a low of 200 to a high of 1000. School #1 has an API of 738 compared to School #2's API ranking of 887. The results suggest that the quality of instructional practices employed at the school play an important role in the school's academic standing.

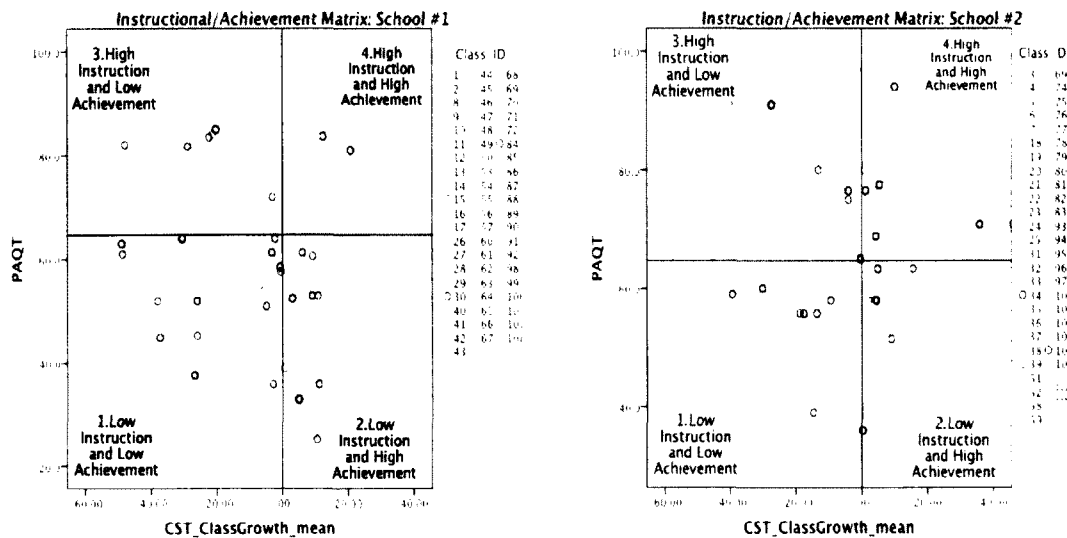


Figure 6. Distribution of instructional quality for School #1 (left) and School #2 (right).

Distribution of PAQT Factors

Up until this point in the analysis, the total PAQT score for each teacher has represented the quality of instruction. The PAQT score represents a composite score including three variables: participation, critical thinking, and academic language. The score for each lesson segment is comprised of the participation score (0–4), the critical thinking score (0–4), and the academic language score (0–4) for a total of 12 points possible for each lesson segment and 120 points possible for the entire lesson comprised of 10 lesson segments. The distribution of scores for each variable that comprises the PAQT score is shown in the figure below for all of the teachers in the sample.

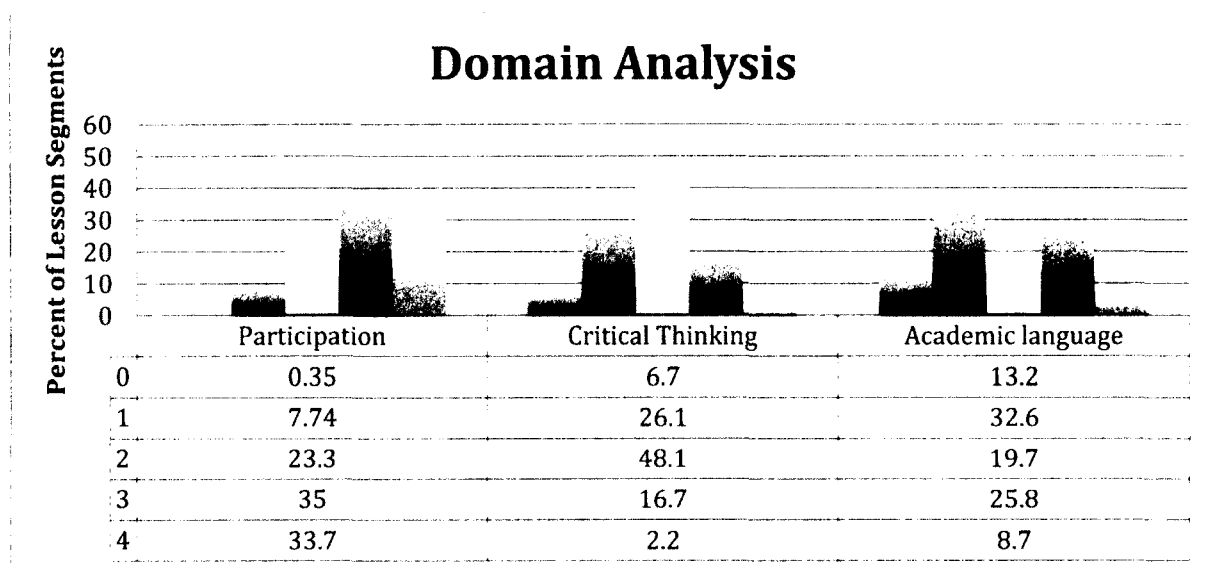


Figure 7. Percent of lesson segments scoring 0–4 for each domain

The graphs show that achieving high scores for critical thinking was the most difficult accomplishment for teachers, as only about 2% of lesson segments required students to engage in extended thinking. This finding indicates that students seldom had opportunities to apply concepts in novel situations, synthesize material, or analyze, critique, or provide justification for their ideas. Opportunities to practice these higher-order thinking skills were absent in the great majority of lessons. It is also interesting to note the prevalence of one-word answers in the academic language category. When students did use academic language (represented by a score of 3 or 4), their communication was limited to a maximum of one sentence and rarely (less than 10% of the time) did the teacher expect students to engage in multi-directional communication where students were in control of their conversations.

A Summary of the Qualitative Study

The qualitative analysis resulted in the identification of eight instructional strategies that were consistently associated with both high instruction and high

achievement and six ineffective practices associated with low instruction and low achievement. The purpose of the qualitative analysis was to describe in detail how high quality instruction is implemented to increase student engagement and learning and to describe ineffective practices so that teachers and administrators can be knowledgeable about teacher actions that negatively impact student engagement and learning. Analysis of the distribution of PAQT variables (participation, critical thinking, and academic language) showed high participation as the most attainable aspect of high-quality instruction and critical thinking as the most challenging component. The next chapter discusses the implications of the study.

CHAPTER FIVE: IMPLICATIONS

The purpose of this study was to investigate a new method for evaluating instructional quality in the classroom. A review of the literature resulted in a working definition of instructional quality: the level at which a teacher facilitates multi-directional interactions with the class that result in authentic cognitive engagement and increased subject matter competence (Cohen, Raudenbush, & Ball, 2003; Goe & Stickler, 2008; Matsumura et al., 2006; Tyler & Boelter, 2008; Westgate & Hughes, 1997). Based on this definition, a theoretical framework was developed for envisioning instructional quality and an instrument was created for measuring instruction in the classroom. The instrument focused on three observable student variables: participation, critical thinking, and academic language. The main goal of the study was to quantify and understand how the quality of a teacher's instruction is related to growth in student achievement.

In the spirit of needing to both measure and understand instructional quality, the study included a quantitative as well as a qualitative component. The quantitative component showed that the variance between teachers in their students' achievement trajectories was substantial, even after accounting for the exogenous variables that impact achievement but are not under the teacher's direct control (i.e., prior year test scores, SES, ethnicity, program participation, behavior, and attendance). Multilevel modeling techniques were used to estimate the effects of classroom instruction on student achievement. In both mathematics and English, the results were positive and significant. For every standard deviation increase in instructional quality (15 points on a scale from 0 to 120), student scores increased 11 points in mathematics and 3.67 points in English. This finding indicates that an average student with a teacher with low instructional quality (one SD below the mean) will score 22 points lower in math and 7.3 points lower in English than that same student would score with a high-quality teacher (one SD above the mean), while all other variables are held constant.

The observation instrument also included a qualitative component, which required the observer to describe the instruction 10 times as the lesson progressed from beginning to end. A matrix was developed that organized each classroom into one of four quadrants: (1) low instruction/low achievement, (2) low instruction/high achievement, (3) high instruction/low achievement, (4) and high instruction/high achievement. Teacher actions in quadrants one and four were analyzed to determine the instructional strategies that resulted in classes with the highest and lowest levels of instruction and achievement.

In classrooms with high instruction and high achievement, eight instructional strategies were identified. These included: (a) immediate engagement, (b) scaffolded

academic dialogue, (c) real-world connections, (d) front loading, (e) differentiated instruction, (f) feedback and conferring, (g) structured reflection, and (h) lesson closure with connected homework. The highest-quality teachers integrated these strategies seamlessly throughout each lesson to maximize student engagement and increase students' subject-matter proficiency.

In classrooms with low instruction and low achievement, six ineffective practices were identified that diminished student engagement and achievement. These included: (a) students waiting with no academic expectations; (b) students copying from the board, texts, or each other; (c) public reprimands for off-task behavior; (d) students being called on one at a time; (e) rapid-fire questioning with one-word answers; and (f) teachers answering their own questions. In these classrooms, it was evident that mastery of the content would be difficult for students. The opportunities students had to make sense of concepts through critical thinking and discussion were limited due to ineffective decision-making on the part of the teacher resulting in low levels of student engagement and achievement.

This study provides further evidence that theoretical models and observation protocols can be developed that accurately describe instructional quality, predict student achievement, and effectively inform the instructional improvement process (Kane, 2010; Schacter & Thum, 2004). The data generated from this study empowers educational leaders to understand instructional observation and the critical role it plays in providing appropriate levels of support to teachers as they build their instructional capacity. These findings have implications in many educational arenas, but especially in the areas of instructional evaluation and professional development as well as in addressing issues of

educational equity. Instructional evaluation can be immediately improved by making observations criterion-referenced, content-focused, unannounced, and consistent. The instructional capacity of districts can be strengthened by aligning professional development to the specific strengths and needs of each teacher, through differentiated support and intervention (Newmann, et al., 2000). Additionally, educational equity can be more confidently assured if leaders use evaluation to inform the critical decisions that impact the academic experiences of students on a daily basis.

Redefining Instructional Evaluation

Instructional evaluation assists teachers in implementing professional practices that enable student success and ensure a proper return on the district's educational investment (Darling-Hammond, 2010). The introduction of the No Child Left Behind Act of 2001 diverted educators' and the public's beliefs away from defining student success in terms student engagement in active learning to simply increasing student test scores (Cuban, 2006; Hargreaves, 2012). Not surprisingly, this shift has impacted the way teachers interpret their role in the classroom in terms of their expectations for both themselves and their students (Elmore, 2009). As such, the system supports the practices of teachers in quadrant two, who obtain increased test results even though their practices do not engage students. Although test scores provide methods for holding all schools accountable for teaching standards-based content and facilitate understanding of the comparative achievement of schools and districts, this study demonstrated that increased test scores alone do not ensure that students have access to high-quality instruction.

In the past, the limitations caused by the difficulties associated with observing instruction meant that little information was available to policymakers, school officials,

and teachers regarding the ways in which instruction may (or may not) transform student understanding. This lack of information was evidenced by Carroll's (1989) study where he was unable to attribute student learning to teachers' instructional practice and Aguirre-Munoz et al.'s (2006) study of the relationship between academic language instruction for English Learners and student achievement, which failed to show a significant effect. At that time, Carroll (1989) attributed the lack of effect to the difficulty of measuring the construct and Aguirre-Munoz et al. (2006) pointed to the lack of variability among teachers in terms of providing strong instruction and differentiated strategies. Both of these issues were addressed in the current study.

Redefining instructional evaluation so that teachers appreciate the process has the potential to increase reflection on and ownership of the impact of their instructional decisions. Evaluation of instruction is to teachers what evaluation of student learning is to students. It is the proven method by which teachers can determine whether or not their actions had the impact they intended. The transition to Common Core State Standards is the profession's attempt to adjust the expectations for student learning, so they more accurately reflect the core purpose of education and the changing educational demands of the 21st century. Because the California Standards curriculum and assessment system was adopted and implemented without a measure of the instructional component, educators were unable to ensure that content was accessible to children through the vehicle of quality instruction. With the addition of valid and reliable evaluation protocols, teachers now have the potential to significantly improve implementation of common core standards-based instruction. Such improvement would require districts to first solidify a common vision of instructional effectiveness. Then, they would need to re-define the

instructional evaluation process to ensure the process is effective. Finally, as teachers develop confidence in the district's ability to ensure effective evaluation, teachers will recognize the value of evaluation as an important mechanism for ensuring instructional quality for every student and improving instruction when it is less than effective.

Adopt a Unified Vision of Instructional Effectiveness

The research is clear about the components of and essential skills needed for high-quality instruction, but many districts struggle to adopt a vision that everyone understands. The theoretical framework upon which the observation protocols in this study were based provides the means for communicating a common vision of quality instruction. It is an integrated theory that describes teaching in terms of content knowledge, pedagogical skills, and classroom culture and measures instruction based on student participation, critical thinking, and academic language. Teachers will be more likely to understand and internalize this vision if they are embedded in a school culture that models and supports it. Instructional leadership requires the skillful facilitation of teacher learning and knowledge construction at all levels of the system, building on the capacity that already exists (Helsing et al., 2008). High-quality school-level leadership, just like high-quality teaching, is dependent on a principal's deep understanding of instruction (knowledge), the ability to provide leadership in a way that builds on teachers' prior knowledge and strengths (pedagogy), and the ability to establish a school culture where everyone is empowered to continually grow and improve (culture).

Modeling high-quality instruction at the school level means that principals are able to lead instruction in the same ways that teachers are able to facilitate learning for students. This nested philosophy (Fink & Resnick, 2001) implies that many of the

effective practices that quadrant four teachers exhibit would be exhibited in the form of instructional leadership by principals. Our eight effective instructional practices can be ramped up to effective leadership practices in the following manner. An effective instructional leader would: (a) continually engage teachers in the instructional improvement process; (b) structure adult learning that fosters serious academic discussions about students' learning; (c) facilitate real-world connections between instruction and our bigger role of impacting students' future lives; (d) front load adult learning to support teacher success in the classroom; (e) differentiate adult learning and scaffold experiences based on individual teacher and group needs; (f) provide meaningful feedback to teachers as they engage in the improvement process; (g) structure venues for active reflection on teaching; and (h) summarize the learning that adults are doing and build meaningful connections to future learning. This vision of engaging instruction, which is modeled at the school level through engaging leadership, would be measured by teachers' participation in learning, their critical thinking about their practice, and their ability to articulate and demonstrate increasingly sophisticated practices that engage students in active learning and achievement. This recommendation implies that school leaders also have access to and actively participate in adult learning activities that increase their own ability to tackle the complex challenges associated with instructional improvement.

Fundamental Principles of Instructional Evaluation

Evaluation in its current form is generally viewed as a negative process that creates a sense of fear and uncertainty. Current models that require site administrators to evaluate all of the teachers in the building as well as manage the school in its entirety do

not support the fundamental principles of effective evaluation. Most quadrant one and two teachers are experienced teachers who have multiple evaluations on record indicating that their instruction is effective. Perfunctory evaluation protocols as they exist currently in some districts result in many principals evaluating teachers as effective even though they do not have the data to substantiate their assessment. When principals have data about instruction that is valid and reliable, they are better able to facilitate the improvement process by recognizing high-quality teachers and dealing directly with ineffectiveness rather than skirting around the issue due to fears about negatively impacting the school's culture. Once a school or district determines quality indicators and agrees on a way of measuring effectiveness, they can select from within their ranks a cadre of instructional evaluators whose responsibility is the improvement of instruction through effective evaluation that meets agreed-upon criteria. This study revealed that effective evaluation is: (a) criterion-referenced, (b) content-focused, (c) unannounced, and (d) consistent.

Criterion-Referenced Observations of Instruction.

Criterion-referenced observations ensure that evaluations are focused on the district's instructional vision and consistent with what we already know about high-quality instruction (Matsumura et al., 2008; Newmann, et al., 1996; Schacter & Thum, 2004). Without criterion-referenced data, it is difficult to convince teachers that the outcomes of the evaluation are worthy of their consideration. Just as students need authentic reasons and informative assessments that engage them in learning in the classroom, teachers need useful, informative, and reliable measures to assess and monitor their own growth and professional competence. Maintaining an atmosphere where

teachers enthusiastically embrace evaluation as a process depends on teachers' positive and productive experiences with evaluation over time. Just as a highly effective math teacher can change a student's attitudes and perceptions about the discipline that may have been previously negative, a highly effective evaluator can do the same for instructional evaluation. When teachers experience evaluation methods that they trust, and they know that the role of the evaluator is to support them in the improvement process, they will be more likely to continually refine their instructional practice as a result of the feedback they receive (Fullan, 2009; Hargreaves & Fullan, 2012; Noffke & Stevenson, 1995; Pfankuch, 1997; Stoll & Fink, 1996).

In this study, students' levels of participation, critical thinking, and academic language served as proxies for teachers' content-area expertise, pedagogical skills, and their ability to facilitate a classroom culture focused on learning. The observation protocol requires the observer to collect data 10 times during the course of a lesson to enable the teacher to "re-see" the lesson's high points and low points in terms participation, critical thinking, and academic language. A graphic depiction of the lesson summarizes the flow of instruction from beginning to end so the teacher can reflect on the lesson segments that result in the highest and lowest levels of student engagement. This very specific, criterion-referenced feedback ensures all teachers receive consistent data that is valid and reliable and facilitates increasingly greater alignment between espoused and enacted instructional outcomes.

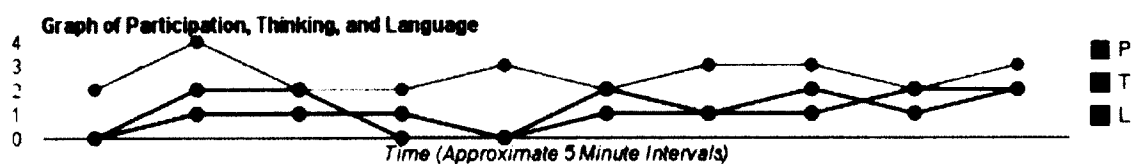


Figure 8: Criterion-referenced feedback summary

Content-Focused Observations of Instruction.

Effective instructional evaluation is dependent on observers having a high degree of content-area expertise. Such expertise allows evaluators to be connoisseurs of instruction who can recognize the qualities that are evident in student–teacher interactions, especially in terms of critical thinking and academic language (Eisner, 2004). This allows evaluators to make teachers cognizant of the aspects of their lessons that are both effective and ineffective as it pertains to the specific parameters of their discipline. Although all school administrators have some degree of classroom experience, many may not have the specific knowledge needed for content-focused evaluations. Even though elementary credentials enable teachers to teach across all disciplines, they do not ensure instructional expertise across all disciplines. For example, it would be very difficult for a principal with a teaching background in physical education or science to effectively evaluate literacy instruction.

In order to ensure effective evaluation, districts need to be creative in the ways that evaluations are scheduled so that administrators can evaluate instruction in the areas in which they have instructional expertise. One way to manage this task would be for evaluators to be assigned teachers to evaluate according to the subject area that was being observed rather than the school in which they are assigned. Such an arrangement would mean that administrators would provide observations with feedback to teachers whom

they do not directly supervise, which has the added bonus of increased objectivity in the evaluation. This increased objectivity adds credence to the evaluation on many levels. If the evaluation shows quadrant one or two level performance, the ensuing discussion would be less likely to generate tension at the school level and the evaluator would be able to serve as a true resource to the teacher in terms of recommended next steps. The principal at the site would have objective data upon which to orchestrate professional learning at the school level and complete final teacher evaluations.

Unannounced Observations of Instruction

It was important in this study that all of the observations in classrooms were unannounced and representative of the typical instruction that happened for students on a daily basis. Even if teachers are capable of high-quality teaching, if they do not perform it on a daily basis, then a planned observation will provide data that is inconsistent with students' actual experience. Because the goal of observational evaluation is to measure the quality of instruction so that correlations can be estimated based on student achievement, it is important that the observations result in accurate measures that reflect teachers' daily practice.

Consistent Observations of Instruction

It is critical that the district have a common method for collecting classroom data so that evaluations are consistent across schools in terms of both quality and quantity (Darling-Hammond & Stanford Center for Opportunity Policy in Education, 2012; Waxman, et al., 2004). Instructional evaluation that is valid and reliable might not be possible when conducted by principals whose primary tasks include an array of school management responsibilities that keep them out of classrooms where learning is expected

to occur (Fink & Resnick, 2001). Many principals struggle because they know they need to spend a significant amount of time in classrooms in order to effectively monitor instructional quality, but few of them are able to give that important role the attention it deserves. Because observations have been so inconsistent in the past, many teachers have lost faith in a principal's ability to even perform the task (Darling-Hammond et al., 2012). One important limitation of this study was that each teacher was observed only once. Ideally, teachers should be observed at least three times during a school year; however, two will also suffice if evaluation protocols ensure year-to-year consistency. Districts can be more proactive to ensure that they dedicate an appropriate amount of time to instructional evaluation, and they can be creative in their structure so that administrators can also learn from the process.

The protocol described in this study has been field-tested in many different contexts, and the reactions from teachers have been overwhelmingly positive. Often, at the end of a debrief conference, teachers are amazed that such explicit feedback is even possible. One teacher said, "Wow. I have never received feedback that has been so detailed." The pilot study included an exit survey where another teacher stated, "When I saw my data, I wanted to figure out how I could make the low points more exciting and engaging for the kids. I would like to get this kind of feedback more often and if I could look at it with my peers, I feel I could really improve my teaching." Another teacher commented that it would be helpful to use the instrument with her peers to assess participation, critical thinking, and academic language after collaborating on planning a lesson. She said, "Then we could really see where we need to tweak our own language to make the instruction stronger." These comments represent an example of the new kinds

of conversations needed regarding the public versus private nature of the teaching profession. Professional knowledge must be made public so that it can be shared, critiqued, and verified (Desimone, 2009), and professional collaboration is effective only if it leads to a deeper understanding of the kinds of conditions and contexts that support and encourage learning (Lieberman & Pointer Mace, 2010). Providing teachers with empirically valid data that allows them to understand the impact that their instruction has on students can be a powerful lever in fostering this level of professional dialogue. Once teachers begin to attribute student gains in performance to their own efforts, the inquiry process and a commitment to continuous improvement are recognized as useful, satisfying, and worth the effort (Gallimore, Ermeling, Saunders, & Goldenberg, 2009).

Building consensus about what high-quality instruction both *is* and *isn't* constitutes the “what” of high-quality evaluation, but not necessarily the “why” or the “how.” The results presented in this study are consistent with previous findings demonstrating that observed instruction falls far below the highest levels described in the proposed framework (Aguirre-Munoz et.al, 2006; Matsumura et al., 2006; Matsumura et al., 2008; Newmann, et al., 1996; Schacter & Thum, 2004). Most teachers did not provide opportunities for students to think critically or engage in rigorous academic discussions, and an over-reliance on direct instruction as the primary teaching method significantly impacted both student engagement and student achievement. Few opportunities for students to engage in rigorous work and discussion might be attributed to the cyclical nature of teaching which results in practitioners applying the instructional techniques which are easiest to deliver, most familiar, and consistent with the way they themselves learn and were taught. The boundary of teacher understanding is constrained

by a teacher's own experience. If teachers do not regularly engage in critical thinking and rigorous conversation in their own lives or in the professional work they do with colleagues, then they may not be cognizant of the need to encourage students to do so. Developing understanding of the rationale for using observation protocols in a consistent manner, which is to help teachers reflect on their instruction and learn new strategies for improving practice, takes place in the context of professional development.

Differentiated Professional Development

As value-added methods for evaluating teacher performance became more prominent after No Child Left Behind (2001), accountability also increased for districts to ensure that if teachers were being held responsible for increasing achievement based on student performance on standardized tests, then districts were also responsible for providing the resources, support, and training needed to foster and improve teachers' instructional competence. As the profession moves to higher standards for student performance as described in the Common Core State Standards, it is critical that teachers be supported in learning how to construct and present tasks that provide multiple opportunities for students to practice higher-order thinking skills. Likewise, as student assessments change to reflect new expectations for learning, the assessment of teaching needs to correspond with those expectations. The data generated from this study demonstrates how professional development can be differentiated to address the specific needs of teachers and how the instructional capacity of schools and districts can be strengthened by more strategic professional development decision-making.

Quadrant-Specific Support and Intervention

An important missing link in the current professional development arena is the connection between teacher practice and the specific professional development opportunities available to teachers (Newmann, et al., 2000). The tools presented in this study provide an observational lens through which to determine a particular teacher's strengths and weaknesses at achieving high levels of instructional performance, and this information can be used to tailor professional development to teachers' very specific needs as learners.

The instructional/achievement matrix described in Chapter four divides teachers into four differentiated learning groups. Quadrant-specific supports are needed that are directly linked to the instruction teachers currently provide for their students. The professional development needs of teachers in quadrant one are very different from the needs of teachers in quadrants two, three, and four. For example, teachers in quadrant one struggle with both content and pedagogy, and the students in their classes are at risk of not learning the content for their grade level. This quadrant represents a major educational liability for both the student and the district and requires a very specialized intervention. Moving teachers out of quadrant one will require them to not only learn new strategies, but also to break ingrained habits of which they may not even be conscious. If teachers are never told that their instruction is ineffective, and how it is ineffective, then like the teachers in quadrant two, they will have no reason to change what they are doing. For example, a teacher who has developed a habit of teaching through the use of recall and rapid-fire questions may succeed in making some students remember the answers and even achieve proficiency on a test. However, such a strategy

will not result in a substantial change in the way the students' brains organize the information. Without the expectation that students explain and justify their answers or apply concepts in unfamiliar contexts, a teacher cannot assume that understanding has been achieved.



Figure 9. Ineffective Practices

Teachers in quadrant two cover the curriculum, but may not be engaging the students in ways that make them excited about what they are learning or that enable them to transfer their knowledge to contexts outside of school. Teachers in this quadrant need opportunities to rediscover their own reasons for becoming a teacher. It was confirmed in a follow-up investigation that all of the teachers in quadrant two had been teaching for at least five years. This finding suggests that it was not their lack of experience that caused them to rely heavily on the textbook or teach to the test, nor was it their lack of experience that caused them to be unconcerned about student engagement. Rather, because they were achieving what the system expected, it is possible that they believed there was no reason for them to change their practice. These teachers may need the added

incentive of high-quality instruction being added as a requirement for an effective formal evaluation. This addition would not require negotiations with the teachers' unions because the evaluation of instruction is already built into most formal evaluation protocols. Some districts have already added a value-added instructional component to teacher pay structures, and as the profession becomes more proficient at generating valid and reliable data, those models are likely to become increasingly more prevalent.

Teachers in quadrant one and two seemed less enthusiastic about the subject matter they taught. They were less likely to use text-based resources strategically to orchestrate learning. Rather, they worked through the text chapters and answered questions sequentially and mechanically, using the questions at the end of each story as a script for their teaching. Similarly, in math, the less-effective teachers focused on worksheets or assignments from the textbook and usually required students to do all of the problems (or, possibly just even or odd problems), rather than focusing on problems or constructing tasks that would yield rich discussions or complex meaning-making. In these instances, the teachers in quadrants one and two resemble line-cooks in a restaurant who are dependent on the chef to know exactly how to measure the ingredients and prepare the meal for the customer's satisfaction or a musician in an orchestra who knows how to read the music and play the instrument, but has not acquired the higher level skills in composition or conducting. Using the test as a single indicator of effective instruction is like checking only the temperature of a meal rather than the taste, the smell, and the presentation; or like assessing only the volume of a musical performance without paying attention to the flow, cadence, or the emotional impact.

Anecdotally, it is interesting to note that none of the teachers in quadrants one or two were new teachers, and in the pilot study, one of the highest scoring teachers on instruction was a student teacher. Her master teacher was a quadrant-four teacher, who was a participant in the current study. This finding suggests that new teachers are highly influenced by the norms they experience in schools and that districts should be more proactive about ensuring that beginning teachers have access to the professional support of colleagues in quadrant four.

Teachers in quadrant three, who taught well but whose students on average did not show growth in achievement, may not have exposed their students to the breadth of the grade-level curriculum or may have failed to make explicit connections between the content learned and the format of particular test questions. These teachers, whose beliefs about instruction are aligned with the research about best practice, might consider an in-depth analysis of the parts of the test on which their students did not perform well to see if their instruction could be adjusted so that it maintained its quality while better preparing their students for proficient performance. Because teachers in quadrant three tend to make strong pedagogical decisions but may struggle to either cover the curriculum that is tested or facilitate connections between what is taught and the way that it is assessed, what these teachers need as learners is different from what teachers in quadrants one and two need. These teachers need to learn strategies for pacing their instruction, better accommodating diverse student needs, and linking instruction and assessment.

Finally, teachers in quadrant four should be leading the work by becoming increasingly more knowledgeable about research and development in both their content

area and in the pedagogical decisions that support heightened engagement in 21st century thinking. All of the quadrant-four classrooms presented evidence of purposeful planning that included a clear objective about what students were expected to learn, a plan for how that learning would be accomplished, and an assessment that informed the teacher about which students achieved understanding. These teachers were not marching through the curriculum as if it were a cookbook for quality teaching. They were using a variety of resources in addition to the standards-based instructional materials to make strategic decisions about instructional tasks. In high-quality classes, there was clear evidence that teachers were critical thinkers themselves who had planned carefully how best to present the material to students and plan for active engagement. Teachers who taught well and whose students earned high test scores demonstrated knowledge about their subject matter and an intrinsic appreciation for academic endeavors.

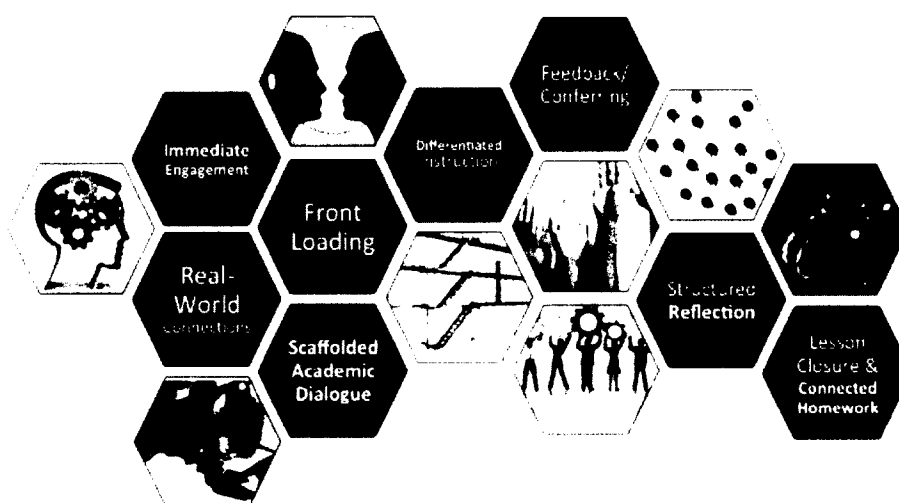


Figure 10. High-quality instructional practices

While we cannot assume that the eight instructional strategies highlighted in this study constitute the full range of options available to high-performing teachers, they do begin to verify the “how” of instructional quality, which is also supported extensively in

the literature (Bowen, 2003; Marzano, 2007; Merrill, 2002; Slavin, Lake, & Groff, 2009; Walqui, 2006; Wang et al., 1990). Undoubtedly, quality practices will become more detailed and extensive as the number of observations and participating schools increase. When teachers receive feedback on the strengths and weaknesses of their lessons, they will need access to resources that assist them in developing new strategies for engaging students. Quadrant four teachers can be instrumental in creating those resources.

Domain-Specific Support and Interventions

The theoretical framework developed for this study emphasized Elmore's (2009) idea that the improvement of instruction can be achieved in one of three ways: (a) by increasing the knowledge and skill of teachers, (b) by increasing the level or complexity of the content, or (c) by significantly altering the role of the student in learning. The data generated from the instructional observations in this study supply the teacher with an in-depth analysis from each domain of the theoretical framework. Levels of participation, critical thinking, and academic language are described numerically and also in narrative form for each of 10 observed lesson segments. From this data, teachers can know exactly what areas of their lessons resulted in highest and lowest levels of student engagement, and professional resources can be acquired based on very specific adult learning goals.

If participation was low, then the teacher may need training in the area of classroom management or in adapting tasks so that students take a more active role. If critical thinking was low, as it was with many of the teachers in this study, professional development can be created that builds teachers' knowledge and experience in creating tasks that are more challenging while, at the same time, supporting students as they learn to practice higher-order thinking. If academic language was low, then professional

development can focus on structuring tasks that scaffold students' use of academic language as they learn and develop conceptual understanding. Because the observation instrument generates more than simply numbers, there are many examples of tasks that result in both high and low levels of participation, critical thinking, and academic language. Analyzing tasks in a comparative fashion helps teachers discover that, by slightly adapting a question or task, the teacher can alter the course of his or her instruction to create more robust student engagement.

For an average teacher who reviews the data generated by an observation and sees that in the majority of lesson segments students focused on *Depth of Knowledge* levels 1 or 2 (Webb, 2005), he or she might be ready to investigate methods for adapting tasks that increase the level of thinking required. Here, appropriate professional development would focus on engaging teachers in tasks that develop their own appreciation for and deep thinking about content. Teachers would be shown models of how effective teachers use and adapt instructional materials to enhance the students' role in thinking, and then work with colleagues to apply new pedagogies to the instruction they deliver to their students. In other words, professional development would provide for teachers what high-quality instruction provides for students. Teachers would have opportunities to develop professional knowledge in an environment that supports their developing understanding of the relationship between content and rigor and pedagogy and academic language practice. When teachers experience for themselves the kind of learning that the new standards require of students, they learn that high expectations and quality interactions are what make the difference and produce the strongest learning outcomes.

Curriculum

Another important consideration for professional development is the use of curriculum resources. The lower effects in English might have been a result of teachers being less likely in general to be following the same curriculum (Matsumura et al., 2008). In English, lessons were much more varied in terms of content and learning expectations than those in math. In the majority of math classes, regardless of quality, teachers were consistent in the use of the structured curriculum as the basis for their lesson's content. The data collected from these observations might have been a stronger indicator of quality since there was less variability in the tasks that teachers employed. The art of teaching in the presence of a structured curriculum had more to do with how the teacher engaged the students, rather than the specific content that was covered. For example, lessons on linear equations were taught in all of the 7th grade math classes, but how it was taught varied depending on the pedagogical decisions teachers made and their effect on student engagement. Some teachers who were following the curriculum but not teaching well (quadrant two), may have achieved high test scores because text-based activities, even if un-engaging, were aligned with what was tested (Hill et al., 2008).

In instances where the teacher was making strategic decisions about lesson content and structure, Matsumura et al. (2008) suggested that the analysis of assignments rather than the teachers' specific actions might serve as a more robust measure of instructional quality. In their study, analyzing tasks designed by teachers tended to capture the critical teaching skills of communicating an objective, providing meaningful practice, and giving feedback to students as the lesson progressed (Matsumura et al., 2008). The highest-performing teachers in this study succeeded in engaging students in

meaningful tasks, freeing themselves up to work with individuals and groups at their specific point of need. These strengths proved difficult for PAQT to capture since the observations were based on data collected from the class as a whole. Generally, once a teacher surpasses a score of 90, the instrument is less able to capture all of the nuances of the most exemplary instruction.

The idea that task creation significantly impacts instructional quality has implications for the curriculum adoption process. Exemplary, quadrant-four teachers should be the ones who evaluate instructional materials and approve district and state adoptions. Most exemplary teachers find state-approved curriculum materials to be significantly under par and adapt lessons to increase participation, critical thinking, and students' use of academic language. Because the current system perpetuates the purchase of inferior curriculum materials, quadrant one and two teachers do not have access to the most engaging instructional tasks, creating a serious differential in students' access to quality.

Communities of Practice

The institutionalization of professional learning communities (PLCs) across the profession has resulted in a contrived collegiality at many sites characterized by mandatory group meetings and formal procedural protocols. While the original intent of the PLC was to increase collaboration and problem-solving among teachers with similar instructional goals (Fullan, 2009; Sergiovanni, 2004), because many practitioners did not understand the rationale for such work, they attended meetings dutifully but did not engage in the thoughtful reflection and collaboration that was needed to improve instruction (Dufour, 2006; King & Kitchener, 2004). PLCs "often degenerated into

meetings about statistics, test results, and short-term fixes rather than longer-term resolutions around deeper goals about instruction and learning” (Hargreaves, 2012, p. 10).

At the same time, current research continues to validate the idea that peer learning among small groups of teachers has the potential to be the most powerful predictor of student achievement over time (Jackson & Bruegmann, 2009). It is the collective experience of teachers as they learn from each other that seems to matter most for improving student achievement. Just as the classroom is the focal point and critical space for understanding and improving student learning, the PLC is the focal point and critical space where teacher learning is evidenced (Osterman & Kottkamp, 2004). PLCs provide space for teachers to collaboratively evaluate the effectiveness of their instruction and, together, plan and make adjustments if student learning falls below the instructional goals that were set (Sergiovanni, 2004).

Communities of practice, on the other hand, provide a more flexible work environment for self-initiated learners. Communities of practice are formed when groups of people who share a concern or a passion for something they do engage either virtually or in person to develop a shared competence and a repertoire of resources (Jones, Fox, Levin, & State Educational Technology Directors Association, 2011). Teachers learn in communities of practice either by reading and studying, attending training, sharing strategies, co-planning lessons, experimenting through action research, lesson study, observing others, or any number of other ways that cause a change in practice to occur. The identifying factor is that these teachers are actively engaged in growing the profession through their own collaborative research, professional writing, co-presentations at conferences, and the co-construction of instructional materials.

It is important to note that the observation protocol described in this study has been tested in three very distinct educational environments. In addition to the middle school implementation described, the instrument was used in a Kenyan secondary school as well as a diverse urban elementary school in the US. In all three contexts, the range of the data was very similar with the lowest performing teachers scoring in the mid to high twenties and the highest performing teachers scoring in the 90s. This suggests the universal nature of high-quality instruction and the applicability of the model in international contexts. Two common themes running across all contexts included generally low expectations for critical thinking and teacher-dominated discourse. The evaluative data produced by the instrument in this study can facilitate improvement in these areas by providing a common framework by which people in multiple locations can compare outcomes and share promising instructional approaches. Because the data is immediately accessible, teachers are equipped to engage in active reflection on teaching and thoughtful and focused collaboration with colleagues at their own site and beyond. Both of these professional activities have been shown to be fundamental attributes of the instructional improvement process (Berry et al., 2010; Boyd & Fales, 1983; Elmore, 1996; Jackson, Kirabo, & Bruegmann, 2009; Schön, 1983).

Professional Efficacy

A teacher's knowledge and beliefs about instruction are embodied in his or her interactions with students (Bowles & Gintis, 2002; Carroll, 1989; Nystrand, 1997). Prioritizing the work in schools so that the interactions between teachers and students take center stage increases the likelihood that schools can capitalize on the existing knowledge base of teachers and integrate new knowledge from research (Barber &

Mourshed, 2007; Darling-Hammond, 2010; Fullan, 2009; Olson, 2003). This step is important because it is the teachers who are ultimately responsible for implementing new practices. A teacher's belief that all students have the capacity to improve their ability as a direct result of the teacher's actions is critical for producing an environment that supports the continual intellectual development of both students and teachers (Boettcher, 2007; Olson, 2003; Porter, 1991; Resnick & Hall, 2003; Schnellert, Butler, & Higginson, 2008). If teachers accept appropriate responsibility and learn to reconfigure tasks until the desired student outcomes are achieved, they will be continually empowered to seek out solutions to students' learning problems (Tucker, Porter, Reinke, Herman, Ivery, Mack, & Jackson, 2005). As soon as they give up this responsibility, however, by blaming students or other entities outside their control, then the belief that what they do matters fails to drive the instructional improvement process.

Although school communities are able to acknowledge in theory that reflection on practice and collaboration with colleagues is important (Argyris & Schön, 1996; DuFour et al., 2006), many teachers in quadrants one and two are still unable or unwilling to engage in active self-reflection and critical inquiry because they remain unconvinced by the data that the outcomes are something over which they have control. Providing assistance for these teachers that allows them to develop a sense of efficacy about their instructional practice would be a monumental achievement for educational leaders (Guskey & Passaro, 1993). Teachers should recognize that, by not attributing deficiencies in their students' learning or engagement to their own lack of knowledge, effort, skill, underlying assumptions, or expectations, they in effect deflect students' lack of progress away from their teaching and onto something outside of themselves (DuFour

et al., 2006). Taking appropriate responsibility for student success or failure assumes that there is more that can be done that is directly under the teacher's control (Bandura, 1993; Tucker, et al., 2005). This responsibility in turn puts the teacher in a more powerful position to discern and develop an action plan to remedy the problem (Argyris & Schön, 1996) and restores ownership of both instruction and student learning to the classroom teacher (Darling-Hammond, 2007; DuFour & Marzano, 2009; Gergen & Thatchenkery, 2004; Weick, 2007).

Schein (2010) asserted that "all forms of learning and change start with some form of dissatisfaction or frustration generated by data that disconfirm our expectations or hopes" (p. 60). Piaget (1993) referred to this process as disequilibrium, a fundamental attribute of learning that is necessary in order to accommodate new schemas. Mezirow's (1997) theory of transformational learning highlights the individual's capacity to use critical reflection and other rational processes to engage in the construction of new knowledge. In the context of education, reflection is viewed as a continuous and spiraling process where educators observe, evaluate, and improve their own teaching practice and in which challenging situations lead to critical analysis and ultimately to new interpretations and awareness (Boyd & Fales, 1983; Osterman & Kottkamp, 2004). Responding productively to defensive behavior or disagreements about why students learned or did not learn enhances the opportunity for collaborative meaning-making (Heifetz, 1999). Kegan (1994) referred to this process as the hidden curriculum, while Argyris (2002) called it double-loop learning, and Schein (2010) called it a learning culture. What is significant about all of these models is that they acknowledge the essential element of adult learning, which is to take responsibility for outcomes and

identify and confront underlying assumptions that obstruct the learning process (Kegan & Lahey, 2009; Mezirow, 1997).

The kind of data that was produced from the tools developed in this study can serve as a powerful force for instructional reflection. Having access to data focused on student responses to teacher actions during instruction facilitates high levels of teacher ownership. If the data is inconsistent with previous assumptions about the impact of their instruction, and individuals or groups can appreciate and hold the anxiety that it produces as a precious resource of valuable information about what is really happening, then they are effectively combating what Kegan and Lahey call the immunity to change (2009). The conflict that is surfaced is a doorway to development and learning (Mezirow, 1997). It allows people to become unstuck from patterns of ineffectiveness and construct meaning of their experiences by forming a public construction of how challenging circumstances will be interpreted and addressed (Ntseane, 2011). This redefinition of what it means to engage in authentic reflective practice drives instructional improvement. This new paradigm creates the conditions under which teachers can see that changes in instruction actually help their students learn better (Cuban, 2103).

The collaborative culture and high level of engaged inquiry that has been described here does not happen automatically, but must instead be nurtured by sustained school- and district-wide efforts that deliberately promote the production of meaningful instructional data and collaboration within and across schools (Osterman & Kottkamp, 2004). Promoting and facilitating reflective group practice is new terrain for leaders who have not had experiences managing and making sense of the complex demands of adult learning (Helsing, et al., 2008; Rooke & Torbert, 2005). If groups can develop the skills

that do not bypass anxiety but instead use it as the basis for productive reasoning in decision-making (Argyris, 2002; Schein, 2010; Trotter, 2006), then DuFour et al.'s (2006) conceptualization of professional learning communities might be more consistently realized.

An important implication here is that the highest-performing teachers (90 and up) should extend their learning by conducting classroom research, co-creating model lessons, and leading communities of practice. These self-initiated, collaborative endeavors represent the highest level of instructional practice and serve as built-in adult learning venues for high-performing teachers. This level of work generates high-quality professional development for proficient teachers as well as the self-generation of strategies and curriculum that could be shared with developing teachers. Developing teachers could observe the expert teachers delivering lessons or they could attend workshops designed by communities of practice as an additional option for professional learning.

Professional efficacy at the leadership level means that administrators also take responsibility for the continual development of teachers in ways that build on teachers' prior knowledge and also address issues when teacher's practices interfere with student learning (Helsing, et al., 2008). Efficacy implies that administrators also take ownership when teachers stagnate or fail to improve when their teaching is not effective. Snow-Gerono (2008) distinguishes between instructional supervision and instructional leadership suggesting that a more procedural understanding of evaluation and professional development results in strict adherence to a set of rules about how evaluation and professional development should be conducted. A more conceptual view implies that

decisions regarding evaluation and professional development are dependent on a range of factors, which can be ascertained by answering a series of questions. For example, if one asks who should lead the evaluation and professional development experience, then one is forced to reflect and consider the teacher's ability to define and direct his or her own learning as opposed to it being a process over which the supervisor has unilateral control. If one inquires about the purpose of the evaluation process, then reflection is required around the tension between evaluation as a summative process versus teacher development as a dynamic and fluid process. Snow-Gerono's (2008) framework illuminates the potential growth that can occur when principals reflect on the tensions between teacher development as a rational practice improved through training in certain techniques, and a more conceptual understanding of evaluation and professional growth that involves reliable data, collaboration, and inquiry. Just as quadrant one and two teachers, whose knowledge of effective practice is externally derived, need to recognize the human and ethical dimensions of teaching and learning, evaluators need to continually reflect on who they are as leaders, what their purpose is, and recognize that they may potentially select from a variety of supports and intervention to create what works best in the situated contexts where teaching and learning occur (Helsing, et al., 2008; Snow-Gerono, 2008). This includes understanding the power relationships that either promote or prevent the creation of a more collegial atmosphere for collaborative and teacher-directed learning while balancing supports for teachers who need more direct supervision. As a profession, it would be wise to heed Snow-Gerono's (2008) advice and consider transcending the steps, stages and procedural understanding of development in favor of more dynamic, cyclical, and multi-dimensional interpretation of the processes

involved in learning so that capacity can be strengthened throughout all levels of the system.

Educational Equity

The findings in this study contribute to the development of increasingly powerful models for measuring instructional quality so that it can be used together with student achievement measures in the assessment of educational equity. Educational equity implies that all students have equal access to and are equally served by high-quality instruction. In this study, 31% of classrooms fell into quadrant one. Because the current evaluation paradigm does not adequately address these deficiencies, the students in those classrooms are at a proven disadvantage.

The ability to uncover and compassionately expose the often unconscious and destructive trends of quadrant one and two teachers are not skills that most school leaders have acquired, which makes it very difficult for systems to enact practices that result in authentic change (Cooper, 2010; Gatimu, 2009; Kegan & Lahey, 2009; Rooke & Torbert, 2005; Tucker, et al., 2005). Because the current incentive system determines teacher effectiveness based on years of service, the amount of professional development attended, units taken, or credentials earned rather than on the quality of instruction, it is not surprising that a teacher would ignore suggestions to change his or her beliefs about the role of the teacher. After all, these teachers have followed all of the rules required of them professionally and, at least for quadrant two teachers, their students are achieving higher test scores than some of their colleagues, who appear more dedicated to improving their practice. When teachers do not have an intrinsic desire to change and the accountability system reinforces an extrinsic focus in terms of pay scales, then the system

overrides the focus on instructional improvement. On the other hand, if instructional outcomes become the focus of determining a teacher's value, then their practice is more likely to become the focus of their attention and effort.

Most quadrant-one teachers seemed unaware that the ineffective practices described in this study were detrimental to student learning. If instructional leaders are not able to bring this data to the teacher's attention and provide resources on how to change the behaviors, then instructional quality will remain at substandard levels, especially for children who are clustered in the most challenging schools (Fink & Resnick, 2001). This study is a call to action for educational leaders who care about closing the achievement gap. Leaders who are serious about addressing these concerns can take three immediate actions: (a) distribute quality teachers evenly across school sites based on student needs, (b) take action to correct ineffective practices, and (c) make cultural proficiency mandatory.

Distribute Quality Teachers Equitably Across Schools Based on Student Needs

Despite the fact that proper assessment and evaluation of teachers is fundamental to successful schools, this key factor in the improvement process is too frequently neglected—due not to the absence of teacher evaluation, but rather to the implementation of poor evaluation practices. Analysis of the instructional/achievement matrices at both schools in this study revealed a disproportionate number of teachers in both quadrants one and four, with more classrooms in quadrant one at School #1 and more classrooms in quadrant four at School #2. Given that the proportion of students most impacted by the achievement gap (African American, Hispanic, English Learners, students with IEPs, and low-income students) is higher at School #1, it would be correct to assume that those

students do not have equal access to high-quality instruction despite all of the teachers being “highly qualified” on paper. Other studies have shown that often the weakest teachers are relegated to teaching the students with the most need (Darling-Hammond, 2007; Hewitt, 2011), as was the case in this study. As the impact of this reality becomes more visible with improved evaluation practices, the profession will need to distribute quadrant-one teachers so that vulnerable subgroups are not repeatedly subjected to poor instruction.

With the advent of tools for discerning each teacher’s strengths, districts should begin to place teachers with students they are able to teach. For example, if certain teachers have a proven track record for teaching African American students (or Hispanic, EL, or students with IEPs) in ways that engage them and strengthen their proficiency, those students should be assigned to those teachers. Because it has been shown that children who have strong teachers will eventually excel, no matter what their background, while children who have even two weak teachers in a row will never recover (Haycock, 2006), issues of equity and access will need to take more of a central role in district and site-based decision-making. If quality is not evenly distributed across schools, then districts cannot claim that all students have access to high-quality teachers.

Acknowledging that teaching the system’s most vulnerable children requires special skill sets which have not been acquired by all teachers has implications for changing teacher pay structures (Waxman, et al., 2004). Higher levels of competence should result in greater compensation. Districts will also need to determine how long a quadrant-one teacher is allowed to remain at that level once supports are in place for addressing the teacher’s deficiencies. If the system allows such a teacher to remain in

classrooms, districts need to be strategic and equitable about teacher assignments so that they are distributed in ways that minimize deleterious effects on students. Closing the achievement gap would be possible if underserved African American and Hispanic students were assigned to four highly effective teachers in a row (Gordon, Kane, & Staiger, 2006), and the data produced by this study makes it possible to identify the teachers most likely to produce those results.

Correct Ineffective Practices

The world's top-performing school systems of Finland and Singapore recognize that individual teachers who are not performing well need to become aware of the ineffective practices that inhibit learning as well as the underlying belief systems that generate those behaviors (Barber & Mourshed, 2007). As stated earlier, teachers cannot be expected to improve their less effective practices unless they are informed of the need to do so by a skilled and reliable instructional evaluator. The instructional evaluator takes on an even more significant role when the teacher is either unable or unwilling to be reflective and analytical about the outcomes produced by his or her instruction.

As the profession accepts the realization that a significant percentage of its teachers are quadrant-one teachers who have not acquired the skills, knowledge, or disposition to teach all students in an equitable manner, systems will need to begin evaluating instruction effectively district-wide so that teachers can be made aware of the concerns and receive assistance. Instructional supervision requires administrators to address these issues directly by insisting that struggling teachers improve so that the children in their care can learn. Quadrant-one and quadrant-two teachers might need to be distributed more strategically across districts to ensure greater access to proficient models

of high-quality instruction. They need to be able to attend training during their workday, which may require a reduced workload, specialized professional development, and opportunities to observe and plan with the support of exemplary teachers. If they are provided with the supports and resources needed to improve, yet remain in quadrant one or two despite the support, then districts need to be more proactive about counseling them out of the teaching profession.

Make Cultural Proficiency Mandatory

Providing all students access to instruction that results in improved academic performance remains a significant challenge for the education profession (James-Wilson & Hancock, 2011). This is evidenced by persistent achievement gaps among underserved populations of students in schools where dropout rates are more than double those of White and Asian students (California Dropout Research Project, 2009; Goe & Stickler, 2008; Hanushek, 2011; Matsumura et. al., 2006; McDougal, Saunders, & Goldenberg, 2007; Schacter & Thum, 2004). These statistics are substantiated by the fixed effects of this study demonstrating the system's inability to address issues of instructional effectiveness, especially for teachers of students who are African American or Hispanic. All other factors being equal, the African American students in this data set achieved on average 18.6 points less in math and 24.5 points less in English compared to their non-African American counterparts. The effects for Hispanic students were roughly equal in both subject areas (about 17 points lower). Picower (2009) pointed out that the profession's failure to address the instructional ineffectiveness of teachers serving urban communities of Color functions to perpetuate the achievement gap. An implicit institutional rule that prevents administrators from accurately assessing instructional

quality is a fear of not wanting to create an environment where people feel inadequate.

“This ‘just be nice’ mentality when it comes to teacher evaluation serves to maintain the status quo while keeping the focus of urban educational failure on students rather than on teachers’ own willful lack of preparation to teach in communities unfamiliar to them” (Picower, 2009). Education reforms have attempted to close these gaps in academic achievement, but despite legal mandates (NCLB, 2001; IDEA, 2004), curricular innovations, teacher training, a wealth of professional literature, advice, and methods about best practices, the gap between most students of Color and their White and Asian counterparts stubbornly persists (Ferguson, 2003; Haycock & Crawford, 2008; Picower, 2009).

Attention to cultural competence is one of the foundations of effective teaching (Pang et al., 2011). Individuals who have developed cultural competence “comprehend, understand, and behave effectively when faced with culturally diverse situations, where assumptions, values, and traditions differ from those traditions with which they are accustomed” (Pang, et al.: 561). The failure to properly read and respond to differences can create serious problems in maintaining a problem-solving mentality when culturally diverse students do not meet teacher expectations in the classroom (Gatimu, 2009; Lindsey, Roberts, & CampbellJones, 2004; Waxman, et al., 1997). By challenging teachers’ stereotypical constructions of under-performing students, the belief is that they will develop greater capacity to identify, empathize, and build relationships based on an authentic sense of caring for the students they serve (Gatimu, 2009; Graff, 2009; James-Wilson & Hancock, 2011; Noddings, 1992; Waxman, et al., 1997; Valenzuela, 1999). The ability to identify and relate to students will assist teachers in creating curriculum

that is relevant to students' lives and interests and actively supports their academic achievement (Ladson-Billings, 1994; Wentzel, 1997).

Conclusion

The quality of instruction in the classroom has been shown to be the most significant and powerful leverage point for school improvement because it is the only thing over which educators have a considerable degree of control (Berry, Daughtrey, & Wieder, 2010; Darling-Hammond, 2000; Haberman, 1992; Rivkin, Hanushek, & Kain, 2001, 2005; Shacter & Thum, 2004; Wenglinsky, 2004). The instrument used in this study attempted to capture and quantify the cacophony of variables that contribute to high-quality instruction and to produce immediate access to a different kind of instructional data that tells how students are engaged or disengaged in the learning process. The world's leading school systems live by the belief that you cannot improve schools without improving instruction (Barber & Mourshed, 2007). Elmore (2000) defined improvement as:

change with direction, sustained over time, that moves entire systems, raising the average level of quality and performance while at the same time decreasing the variation among units, and engaging people in analysis and understanding of why some actions seem to work and others don't. (p. 13)

Many districts assume that, because they have a teacher assigned to a position, teaching and learning are being accomplished. However, this has been shown to not always be the case, and it is a risky assumption to make if one is concerned about all students having

access to strong instruction (Hargreaves, 2012). Schacter and Thum (2004) agree that instructional evaluations, in conjunction with student achievement gains, comprise a successful accountability system. Integrating an instructional component into a framework for understanding teaching, learning, and educational equity represents a step toward a more complex understanding of the role of evaluation for both teacher development and student learning (Snow-Gerono, 2008). Effective, criterion-referenced evaluation practices ensure that quality and performance increase while variation among classrooms decreases, which levels the playing field for all students. Effective evaluation and professional development results in all teachers across all quadrants getting the differentiated support they need to continually improve their practice. Now that we have a variety of tools to identify the quality of instruction as well as tools to measure, monitor, and develop cultural proficiency (James-Wilson & Hancock, 2011; Lindsey et al., 2004), it is our responsibility to use them to dramatically improve learning for students most dependent on their education as a means of achieving success.

References

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, 25(1), 95-135.
- Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., Raths, J., & Wittrock, M. (Eds.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Addison Wesley Longman, Inc.
- Argyris, C. (1990). *Overcoming organizational defenses: Facilitating organizational learning*. Needham Heights, MA US: Allyn & Bacon.
- Argyris, C. (1993). *Knowledge for Action: A Guide to Overcoming Barriers to Organizational Change*.
- Argyris, C. (2002). Double-loop learning, teaching, and research. *Academy Of Management Learning & Education*, 1(2), 206-218.
- Argyris, C. and Schön, D. (1996). *Organizational learning II: Theory, method, practice*. Addison-Wesley Publishing Company, Reading, MA.
- Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: autonomy-enhancing and suppressing teacher behaviors predicting students' engagement in schoolwork. *British Journal of Educational Psychology*, 72(2), 261-278.
- Ball, D. L., & Cohen, D. (1999). Developing practice, developing practitioners. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3-32). San Francisco, CA: Jossey Bass Publishers.
- Ball, D., & Rowan, B. (2004). Introduction: Measuring Instruction. *Elementary School Journal* (pp. 3-10). University of Chicago Press.
- Bakhtin, M., Holquist, M., & Emerson, C. (1981). *The dialogic imagination: Four essays by M. M. Bakhtin*. Austin: U of Texas P.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117-148.
- Barber, M., & Mourshed, M. (2007). *The McKinsey Report: How the world's best performing school systems come out on top*. London: McKinsey & Company.

- Berry, B., Daughtrey, A., and Wieder, A. (2010). *Teacher effectiveness: The conditions that matter most and a look to the future*. Hillsborough, NC: Center for Teaching Quality.
- Bertrand, Y. (2003). *Contemporary theories and practice in education* (2nd ed.). United States of America: Atwood Pub.
- Bloom, Benjamin S. (Ed). (1956). *Taxonomy of educational objectives handbook 1: Cognitive domain*. New York: Longman, Green & Co.
- Bloom, B. S. (1978). New views of the learner: Implications for instruction and curriculum. *Educational Leadership*, 35(7), 563.
- Boettcher, J. V. (2007). Ten core principles for designing effective learning environments: Insights from brain research and pedagogical theory. *Innovate: Journal of Online Education*, 3(3).
- Boser, U., & Center for American, P. (2011). *Return on educational investment: A district-by-district evaluation of U.S. educational productivity*. Center For American Progress.
- Bowen, E. R. (2003). Student engagement and its relation to quality work design: A review of the literature. *Action Research Exchange*, 2(1).
- Bowles, S., & Gintis, H. (2002). Schooling in Capitalist America Revisited. *Sociology of Education*, 75(1), 1-18.
- Boyd, E. M., & Fales, A. W. (1983). Reflective learning: Key to learning from experience. *Journal Of Humanistic Psychology*, 23(2), 99.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Bruner, Jerome S. (1966). *Toward a Theory of Instruction*. (pp. 49–53). Cambridge MA: Harvard University Press.
- California Department of Education. (2012, June 27). *Release # 12-65*. Retrieved from: <http://www.cde.ca.gov/nr/ne/yr12/yr12rel65.asp>
- California Teacher's Association (2012). CTA Teacher Evaluation Framework.
- CampbellJones, F., CampbellJones, B., & Lindsey, R. B. (2010). *The cultural proficiency journey: Moving beyond ethical barriers toward profound school change*. Thousand Oaks, CA US: Corwin Press.
- Caracelli, V. J., & Greene, J. C. (1997). Crafting Mixed-Method Evaluation Designs. *New Directions for Evaluation*, (74), 19-32.

- Carroll, J. (1963). A model of school learning. *Teachers College Record*, 64, 723-733.
- Carroll, J. B. (1989). The Carroll Model: A 25-Year Retrospective and Prospective View. *Educational Researcher*, 18(1), 26-31.
- Cazden, C. B. (2001). *Classroom discourse: The language of teaching and learning* (2nd ed.). Portsmouth, NH: Heinemann.
- Chelmsky, E. (1998). The role of experience in formulating theories of evaluation practice. *American Journal of Evaluation*, 19(1), 35-56.
- Clark, R. E., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning: The case for fully guided instruction. *American Educator*, 36(1), 6-11.
- Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement*. Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania (CPRE RR-43).
- Cohen, D. K., Raudenbush, S. W., & Ball, D. (2003). Resources, Instruction, and Research. *Educational Evaluation & Policy Analysis*, 25(2), 119.
- Collins, A. (1990, March). *Transforming the assessment of teachers: Notes on a theory of assessment for the 21st century*. Paper presented at the annual meeting of the National Catholic Education Association, Boston, MA.
- Corno, L., & Mandinach, E. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist*, 18, 88-108.
- Costa, A. L., Kallick, B., & Association for Supervision and Curriculum, D. (2008). *Learning and leading with habits of mind: 16 essential characteristics for success*. Association for Supervision and Curriculum Development.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. Second edition.
- Cuban, L. (2006). Getting Past Futile Pedagogical Wars. *Phi Delta Kappan*, 87(10), 793-795.
- Cuban, L. (2013). Why so many structural changes in schools and so little reform in teaching practice? *Journal of Educational Administration*, 51(2), 109-125. doi:10.1108/09578231311304661
- Daggett, W. (2002). *The rigor and relevance handbook*. New York: International Center for Leadership in Education.
- Daggett, W. R. (2004) *Reforming American High Schools: Why, what, and how*. Rexford, NY: International Center for Leadership in Education.

- Daley, G., Kim, L., & National Institute for Excellence in Teaching, (2010). *A teacher evaluation system that works: Working paper*. National Institute For Excellence In Teaching.
- Danielson, C. (1996). *Enhancing Professional Practice: A Framework for Teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson, C. (2007). *Enhancing Professional Practice: A Framework for Teaching* (2nd Ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. The Jossey-Bass Education Series.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1).
- Darling-Hammond, L. & Sykes, G. (2003). Wanted: A national teacher supply policy for education: The right way to meet the 'highly qualified teacher' challenge. *Education Policy Analysis Archives*, 11(33).
- Darling-Hammond, L. (2007). Recognizing and enhancing teacher effectiveness: A policymaker's guide. In L. Darling-Hammond and C. D. Prince (Eds.), *Strengthening Teacher Quality in High-Need Schools: Policy and Practice*. Washington, DC: The Council of Chief State School Officers.
- Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. NY: Teachers College Press.
- Darling-Hammond, L. (2011). Restoring our schools: The quest for equity in the United States. *Education Canada*, 51(5), 14-18.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8-15.
- Darling-Hammond, L., & Stanford Center for Opportunity Policy in Education, (2012). *Creating a comprehensive system for evaluating and supporting effective teaching*. Stanford Center For Opportunity Policy In Education.
- Davis, H. A. (2006). Exploring the contexts of relationship quality between middle school students and teachers. *Elementary School Journal*, 106(3), 193-223.
- Delpit, L. (1988). *The silenced dialogue: power and pedagogy in educating other people's children*. Harvard Educational Review.

- Desimone, L. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181-199.
- Dewey, J. (1910). *How we think*. Lexington, MA US: D C Heath. doi:10.1037/10903-000
- Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. New York, NY US: Macmillan Publishing.
- Dewey, J. (2010). The need for a philosophy of education (1934). *Schools: Studies in Education*, 7(2), 244-245.
- Donato, R., & Lazerson, M. (2000). New Directions in American Educational History: Problems and Prospects. *Educational Researcher*, 29(8), 4-15.
- Dooner, A., Mandzuk, D., Obendoerfer, P., Babiuk, G., Cerqueira-Vassallo, G., Force, V., & ... Roy, D. (2010). Examining student engagement and authority: Developing learning relationships in the middle grades. *Middle School Journal*, 41(4), 28-35.
- DuFour, R., DuFour, R., Eaker, R., & Many, T. (2006). *Learning by doing: A handbook for professional learning communities at work*. Bloomington, IN: Solution Tree.
- DuFour, R., & Marzano, R. J. (2009). High-level strategies for principal leadership. *Educational Leadership*, 66(5), 62-68.
- Dwyer, C. (1998). Psychometrics of Praxis III: Classroom performance assessments. *Journal of Personnel Evaluation In Education*, 12(2), 163-187.
- Echevarria, J., Vogt, M. E., & Short, D. (2000). *Making content comprehensible to English language learners: The SIOP model*. Boston: Allyn & Bacon.
- Edwards, A. D., & Westgate, D. P. G. (1994). *Investigating classroom talk* (2nd ed.). London: The Falmer Press.
- Eisner, E. (2004). The roots of connoisseurship and criticism: A personal journey. In M. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences*. Thousand Oaks, CA: Sage.
- Elmore, R. F. (1996). Getting to scale with successful educational practices. *Harvard Educational Review*, 66(1), 1-26.
- Elmore, R. (2000). *Building a new structure for school leadership*. Washington, DC: The Albert Shanker Institute.
- Elmore, R. F. (2002). *Bridging the gap between standards and achievement: The imperative for professional development in education*.

- Elmore, R. F. (2009). School improvement and the reduction of poverty. In M. Bane, R. Zenteno (Eds.), *Poverty and Poverty Alleviation Strategies in North America* (pp. 165-199). David Rockefeller Center Series on Latin American Studies. Cambridge: Harvard University, Press.
- Ferguson, R. F. (2003). Teachers' perceptions and expectations and the Black-White test score gap. *Urban Education*, 38(4), 460-507.
- Fink, E., & Resnick, L. B. (2001). Developing principals as instructional leaders. *Phi Delta Kappan*, 82(8), 598-606.
- Flynt, E., & Brozo, W. G. (2009). It's all about the teacher. *Reading Teacher*, 62(6), 536-538.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review Of Educational Research*, 74(1), 59-109.
- Freire, P. (1970). *Pedagogy of the Oppressed*. New York; Harmondsworth, England: Herder; Penguin.
- Frase, L. E., & Streshly, W. (1994). Lack of accuracy, feedback, and commitment in teacher evaluation. *Journal of Personnel Evaluation In Education*, 8(1), 47-57.
- Fullan, M. (2009). Large-scale reform comes of age. *Journal of Educational Change*, 10(2/3), 101-113. doi:10.1007/s10833-009-9108-z
- Gallimore, R., Ermeling, B. A., Saunders, W. M., & Goldenberg, C. (2009). *Moving the learning of teaching closer to practice: Teacher education implications of school-based inquiry teams*.
- Gardner, Howard. (1983) *Frames of mind: The theory of multiple intelligences*. New York: Basic Books, Inc.
- Gatimu, M. M. (2009). Undermining critical consciousness unconsciously: Restoring hope in the multicultural education idea. *Journal of Educational Change*, 10(1), 47-61. doi:10.1007/s10833-008-9087-5
- Glaser, R. (1976). Components of a psychology of instruction: Toward a science of design. *Review of Educational Research*.
- Goe, L., & National Comprehensive Center for Teacher, Q. (2007). *The link between teacher quality and student outcomes: A research synthesis*. National Comprehensive Center For Teacher Quality.

- Goe, L., Bell, C., Little, O., & National Comprehensive Center for Teacher, Q. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. National Comprehensive Center For Teacher Quality.
- Goe, L., Stickler, L. M., & National Comprehensive Center for Teacher, Q. (2008). Teacher quality and student achievement: Making the most of recent research. *TQ Research & Policy Brief*. National Comprehensive Center For Teacher Quality.
- Goldhaber, D. (2002). The mystery of good teaching: Surveying the evidence on student achievement and teachers' characteristics. *Education Next* 2(1): 50–55.
- Goldhaber, D., & Anthony, E. (2004). *Can teacher quality be effectively assessed?* Center on Reinventing Public Education, Evans School of Public Affairs, University of Washington and the Urban Institute.
- Gordon, R., Kane, T. J., & Staiger, D. O. (2006). *Identifying effective teachers using performance on the job*. Washington, DC: Brookings Institution.
- Graff, N. (2009). Classroom talk: co-constructing a 'difficult student'. *Educational Research*, 51(4), 439-454. doi:10.1080/00131880903354782
- Grossman, P. L., Stodolsky, S. S., & Knapp, M. (2004). *Making subject matter part of the equation: The intersection of policy and content*. Occasional paper (Document 0-04-1), Center for the Study of Teaching and Policy, University of Washington, Seattle.
- Guskey, T. R., & Passaro, P. (1993). *Teacher efficacy: A study of construct dimensions*.
- Haberman, M. (1992). The role of the classroom teacher as a curriculum leader. *NASSP Bulletin*, 76(547), 11-19.
- Haertel, G. D., Illinois Univ., C. s., & And, O. (1980). *Psychological models of educational performance: A theoretical synthesis of constructs*.
- Halliday, M. K. (1993). Towards a language-based theory of learning. *Linguistics And Education: An International Research Journal*, 5(2), 93-116.
- Hampel, R. (1996). Tinkering toward utopia: A century of public school reform. *History of Education Quarterly*, 36(4), 473.
- Hancock, V., & Betts, F. (2002). Back to the future: Preparing learners for academic success in 2004. *Learning And Leading With Technology*, 29(7), 10-27.
- Hanushek, E. A. (2011). Valuing teachers: How much is a good teacher worth? *Education Next*, 11(3), 41-45.

- Hargreaves, A., Shirley, D., & National Staff Development, C. (2009). *The fourth way: The inspiring future for educational change*. Corwin.
- Hargreaves, A. (2012). Singapore: The fourth way in action. *Educational Research For Policy And Practice*, 11(1), 7-17.
- Hargreaves, A., & Fullan, M. (2012). *Professional capital: Transforming teaching in every school*. Teachers College Press.
- Harris, B. (2011). Battling boredom: 99 strategies to spark student engagement. *Eye on Education*.
- Harris, L. R. (2008). A phenomenographic investigation of teacher conceptions of student engagement in learning. *Australian Educational Researcher*: 35(1), 57-79.
- Hatano, G. & Osuro, Y. (2003). Commentary: Reconceptualizing school learning using insight from expertise research. *Educational Researcher*, 32, 26-29.
- Hattie, J. (2009). *Visible learning*. London: Routledge.
- Haycock, K. & Crawford, C. (2008). Closing the teacher quality gap. *Educational Leadership*, 00131784, Apr2008, Vol. 65, Issue 7
- Heck, Ronald H. (2007). Examining the relationship between teacher quality as an organizational property of schools and students' achievement and growth rates. *Educational Administration Quarterly* 2007 43: 399.
- Heck, R. H., & Hallinger, P. (2009). Assessing the contribution of distributed leadership to school improvement and growth in math achievement. *American Educational Research Journal*, 46(3), 659-689.
- Heck, R. H., & Moriyama, K. (2010). Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: A regression discontinuity approach. *School Effectiveness And School Improvement*, 21(4), 377-408.
- Heck, R. H., Scott, T. L., & Tabata, L. N. (2010). *Multilevel and longitudinal modeling with IBM SPSS*. New York, NY: Routledge.
- Helsing, D., Howell, A., Kegan, R., & Lahey, L. (2008). *Putting the "development" in professional development: Understanding and overturning educational leaders' immunities to change*. *Harvard Educational Review*, 78(3), 437-465.
- Henig, J. R., & Stone, C. N. (2008). Rethinking school reform: The distractions of dogma and the potential for a new politics of progressive pragmatism. *American Journal of Education*, 114(3), 191-218.

- Hernández Finch, M. E. (2012). Special considerations with response to intervention and instruction for students with diverse backgrounds. *Psychology in the Schools*, 49(3), 285-296. doi:10.1002/pits.21597
- Hess, K. K., Jones, B. S., Carlock, D., & Walkup, J. R. (2009). *Cognitive rigor: Blending the strengths of Bloom's taxonomy and Webb's depth of knowledge to enhance classroom-level processes*. Online Submission.
- Hewitt, D. T. (2011). Reauthorize, revise, and remember: Refocusing the no child left behind act to fulfill brown's promise. *Yale Law & Policy Review*, 30(1), 169-194.
- Hidi, S., & Renninger, K. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111-127.
- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371-404). Charlotte, NC: Information Age Publishing.
- Hill, H. C., & Ball, D. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education*, 35(5), 330-351.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42, 371-406.
- Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition & Instruction*, 26(4), 430-511.
- Hox, J. J. (2010). *Applied multilevel analysis* (2nd Edition). New York: Routledge.
- IBM Corp. (2011). *IBM SPSS Statistics for MAC*, Version 20.0. Armonk, NY: IBM Corp.
- International Center for Leadership in Education. (2002). *Rigor and relevance handbook*. Rexford, NY: Author.
- Jackson, C. Kirabo, and Elias Bruegmann. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics* 1(4).
- Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101-136.

- James-Wilson, S., & Hancock, M. (2011). 5 stages on the path to equity: framework challenges urban teachers' deficit thinking. *Journal of Staff Development*, 32(3), 26-28.
- Jones, R., Fox, C., Levin, D., & State Educational Technology Directors Association (2011). National Educational Technology Trends: 2011. *Transforming education to ensure all students are successful in the 21st century*. State Educational Technology Directors Association.
- Junker, B. W., Matsumura, L. C., Crosson, A., Wolf, M. K., Levison, A., Wiesberg, J., et al. (2006). *Overview of the Instructional Quality Assessment. CSE Technical Report #67*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, 46(3), 587-613.
- Kegan, R. (1994). *In over our heads: the mental demands of modern life*. Cambridge, Mass: Harvard University Press.
- Kegan, R., & Lahey, L. (2009). *Immunity to change: How to overcome it and unlock the potential in yourself and your organization*. Harvard Business School Press Books, 1.
- Kennedy, M. M. 2008. Sorting out teacher quality. *Phi Delta Kappan* 90(5).
- Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34-70.
- King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, 31(2), 338-368.
- King, P. M., & Kitchener, K. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist*, 39(1), 5-18.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41, 75-86.
- Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34-70.

- Kohn, A. (1999). *The schools our children deserve: Moving beyond traditional classrooms and "tougher standards"*. New York: Houghton Mifflin.
- Kong, S., & Hoare, P. (2011). Cognitive content engagement in content-based language teaching. *Language Teaching Research*, 15(3), 307-324.
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Neubrand, M., Blum, W., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, 100(3), 716-725.
- Kyriakides, L., & Campbell, R. J. (2003). Teacher evaluation in Cyprus: Some conceptual and methodological issues arising from teacher and school effectiveness research. *Journal of Personnel Evaluation In Education*, 21.
- Kyriakides, L., Demetriou, D., & Charalambous, C. (2006). Generating criteria for evaluating teachers through teacher effectiveness research. *Educational Research*, 48(1), 1-20.
- Ladwig, J. (2007). Modeling pedagogy in Australian school reform. *Pedagogies*, 2(2), 57-76.
- Ladwig, J. (2009). Working backwards towards curriculum: On the curricular implications of Quality Teaching. *Curriculum Journal*, 20(3), 271-286.
- Lieberman, A., & Pointer Mace, D. (2010). Making practice public: teacher learning in the 21st century. *Journal of Teacher Education*, 61(1-2), 77+.
- Lindsey, R. B., Roberts, L. M., & CampbellJones, F. (2004). *The culturally proficient school: An implementation guide for school leaders*. Corwin Press.
- Linnenbrink, E. A., & Pintrich, P. R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading & Writing Quarterly*, 19(2), 119.
- Lounsbury, J. H., Clark, D. C., & National Association of Secondary School Principals, R. A. (1990). *Inside grade eight: From apathy to excitement*.
- Marshall, K. (2012a). Fine-tuning teacher evaluation. *Educational Leadership*, 70(3), 50-53.
- Marshall, K. (2012b). Let's cancel the dog-and-pony show. *Phi Delta Kappan*, 94(3), 19-23.
- Marzano, R. J., & Association for Supervision and Curriculum, D. (2007). *The art and science of teaching: A comprehensive framework for effective instruction*. Association for Supervision and Curriculum Development.

- Matsumura, L., Slater, S., Junker, B., Peterson, M., Boston, M., Steele, M., & ... National Center for Research on Evaluation, S. A. (2006). *Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the instructional quality assessment. CSE technical report 681*. National Center For Research On Evaluation, Standards, And Student Testing (CRESST).
- Matsumura, L., Garnier, H. E., Slater, S., & Boston, M. D. (2008). Toward measuring instructional interactions "at-scale". *Educational Assessment*, 13(4), 267-300.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND.
- McDonnell, L. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation & Policy Analysis*, 17(3), 305-322.
- McDougall, D., Saunders, W. M., & Goldenberg, C. (2007). Inside the black box of school reform: Explaining the how and why of change at getting results schools. *International Journal of Disability, Development & Education*, 54(1), 51-89. doi:10.1080/10349120601149755
- McLaughlin, M. W., Shepard, L. A., & National Academy of Education, S. A. (1995). *Improving education through standards-based reform. A Report by the National Academy of Education Panel on Standards-Based Education Reform*.
- Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, MA: Harvard University Press.
- Mercer, N. (1995). *The guided construction of classroom knowledge*. Clevedon: Multilingual Matters.
- Merrill, M. (2002). First principles of instruction. *Educational Technology Research And Development*, 50(3), 43-59.
- Mertens, D. M. (1999). Inclusive evaluation: Implications of transformative theory for evaluation. *American Journal of Evaluation*, 20(1), 1.
- Mezirow, J. (1997). Transformative learning: Theory to practice. *New Directions For Adult & Continuing Education*, (74), 5.
- Milanowski, A. T. (2011). *Validity research on teacher evaluation systems based on the framework for teaching*. Online Submission.
- Nasstrom, G. (2009). Interpretation of standards with bloom's revised taxonomy: a comparison of teachers and assessment experts. *International Journal of Research & Method In Education*, 32(1), 39-51.

- National Board Resource Center at Stanford, U. (2010). *A quality teacher in every classroom: Creating a teacher evaluation system that works for California. Perspectives from accomplished California teachers*. National Board Resource Center At Stanford University.
- National Board for Professional Teaching Standards, A. A. (2002). *What teachers should know and be able to do*.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform: a report to the nation and the secretary of education united states department of education*.
- Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education*, 104(280-312).
- Newmann, F. M., King, M., & Youngs, P. (2000). Professional development that addresses school capacity: Lessons from urban elementary schools. *American Journal of Education*, 108(4), 259.
- Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: an exploration of stability across models and contexts. *Education Policy Analysis Archives*, 18(23).
- Newton, X. A., & Llosa, L. (2010). Toward a more nuanced approach to program effectiveness assessment: hierarchical linear models in k-12 program evaluation. *American Journal of Evaluation*, 31(2), 162-179. doi:10.1177/1098214010363022
- Nystrand, M. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York: Teachers College Press.
- Olson, D. R. (2003). *Psychological theory and educational reform: How school remakes mind and society*. Cambridge: Cambridge University Press.
- Osterman, K. F., & Kottkamp, R. B. (2004). *Reflective practice for educators: Professional development to improve student learning* (2nd ed.). Thousand Oaks, CA US: Corwin Press.
- Palmer, P. J. (1990). Good teaching: a matter of living the mystery. *Change*, 22(10-16).
- Palmer, P. J. (1998). *The courage to teach: exploring the inner landscape of a teacher's life*. San Francisco: Jossey-Bass Pub.
- Pang, V., Stein, R., Gomez, M., Matas, A., & Shimogori, Y. (2011). Cultural competencies: essential elements of caring-centered multicultural education. *Action In Teacher Education*, 33(5/6), 560-574.

- Patton, M. Q. (2008). *Utilization focused evaluation* (4th ed.). Thousand Oaks, CA: Sage Publications.
- Pearson, P. (2004). The reading wars. *Educational Policy*, 18(1), 216-252.
- Perkins, D., Jay, E., & Tishman, S. (1993). New conceptions of thinking: From ontology to education. *Educational Psychologist*, 28(1), 67.
- Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74(4), 525-556.
- Piaget, J. (1993). The epigenetic system and the development of cognitive functions. In M. Johnson (Ed.), *Brain development and cognition: A reader* (pp. 31-38). Malden: Blackwell Publishing.
- Picower B. The unexamined Whiteness of teaching: How White teachers maintain and enact dominant racial ideologies. *Race, Ethnicity & Education* [serial online]. July 2009;12(2):197-215.
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33-40.
- Porter, A. (1991). Creating a system of school process indicators. *Educational Evaluation and Policy Analysis* 13(1), 13-29.
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3-14.
- Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models for social and behavioral research: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Ravitch, D. (2000). *Left back: A century of failed school reforms*.
- Regnier, P. (1994). The illusion of technique and the intellectual life of schools. *Phi Delta Kappan*, 7682-83.
- Resnick, L., & Glennan, J., T. K. (2002). Leadership for learning: A theory of action for urban school districts. In A. M. Hightower, M. S. Knapp, J. A. Marsh & M. W. McLaughlin (Eds.), *School districts and instructional renewal: Critical issues in educational leadership* (pp. 160-172). New York: Teachers College Press.
- Resnick, L. B. & Hall, M. W. (2001). *The Principles of Learning: Study tools for educators* [CD-ROM, version 2.0]. University of Pittsburgh, Learning Research and Development Center, Institute for Learning.

- Rivkin, S. G., E. A. Hanushek, and J. F. Kain. 2005. Teachers, schools, and academic achievement. *Econometrica* 73(2): 417–458.
- Rogers, J., Bertrand, M., Perez, W., & UCLA, I. (2012). Finding common ground in education values: Influential Californians speak on the purpose of public education. *UCLA IDEA*.
- Rooke, D., & Torbert, W. (2005). Seven transformations of leadership. *Harvard Business Review*, 83(4), 66-76.
- Rouse, C., & Barrow, L. (2006). U.S. elementary and secondary schools: Equalizing opportunity or replicating the status quo? *Future of Children*, 16(2), 99-123.
- Rumberger, R. W., & Lim, S. (2008). *Why students drop out of school: A review of 25 Years of research*. Santa Barbara, CA: California Dropout Research Project.
- Sanders, W. L. & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.
- Sanders, W. L. (1998). Value-added assessment. *The School Administrator*, 5,5 (11).
- Sawyer, R. (2006). Introduction: The New Science of Learning. In R. Sawyer (Ed.). *The Cambridge handbook of: The learning sciences* (pp. 1-16). New York, NY US: Cambridge University Press.
- Schacter, J., & Thum, Y. M. (2004). Paying for high- and low-quality teaching. *Economics of Education Review*, 23, 411–430.
- Schein, E. H. (2010). *Organizational culture and leadership*. San Francisco: Jossey-Bass.
- Schnellert, L. M., Butler, D. L., & Higginson, S. K. (2008). Co-constructors of data, co-constructors of meaning: Teacher professional development in an age of accountability. *Teaching and Teacher Education: An International Journal of Research and Studies*, 24(3), 725-750.
- Schön, D. (1983). *The reflective practitioner*. New York: Basic Books.
- Sergiovanni, T. J. (2004). Collaborative cultures and communities of practice. *Principal Leadership*, 5(1), 48-52.
- Shernoff, D. J., & Csikszentmihalyi, M. (2009). Flow in schools: Cultivating engaged learners and optimal learning environments. In R. Gilman, E. Huebner, M. J. Furlong (Eds.), *Handbook of positive psychology in schools* (pp. 131-145). New York, NY US: Routledge/Taylor & Francis Group.

- Schraw, G., & Robinson, D. H. (2011). *Assessment of higher order thinking skills. Current perspectives on cognition, learning and instruction*. IAP - Information Age Publishing, Inc.
- Shinkfield, A. J., Stufflebeam, D. L., & Center for Research in Educational Accountability and Teacher Evaluation (CREATE), K. I. (1995). Teacher evaluation: Guide to effective practice. *Evaluation in Education and Human Services*. Kalamazoo, MI: CREATE.
- Shulman, L. S. (1986). Those who understand: A conception of teacher knowledge. *American Educator*, 10(1), 9-15,43-44.
- Silver, E. A., Ghouseini, H., Gosen, D., Charalambous, C., & Strawhun, B. (2005). Moving from rhetoric to praxis: Issues faced by teachers in having students consider multiple solutions for problems in the mathematics classroom. *The Journal of Mathematical Behavior*, 24(3-4), 287-301.
doi:10.1016/j.jmathb.2005.09.009
- Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research*, 79(2), 839-911.
- Smith, K. S., & Geller, C. (2004). Essential principles of effective mathematics instruction: Methods to reach all students. *Preventing school failure*, 48(4), 22-29.
- Smylie, M. (1995). Teacher learning in the workplace: Implications for school reform. In T.R. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms and practices* (pp.92-113). New York: Teachers College Press.
- Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, 2(4), 325-344.
- Snow-Gerono, J. L. (2008). Locating supervision--a reflective framework for negotiating tensions within conceptual and procedural foci for teacher development. *Teaching And Teacher Education: An International Journal Of Research And Studies*, 24(6), 1502-1515.
- Sofa, R. (2008). Beyond NCLB and AYP: One superintendent's experience of school district reform. *Harvard Educational Review*: 78; 2.
- Soter, A. O., Wilkinson, I. A., Murphy, P., Rudge, L., Reninger, K., & Edwards, M. (2008). What the discourse tells us: Talk and indicators of high-level comprehension. *International Journal of Educational Research*, 47(6), 372-391.
- Stake, R. (2004). Stake and responsive evaluation. In M. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences*. Thousand Oaks, CA: Sage.

- Sternberg, R. J. (2003). What is an “expert student”? *Educational Researcher*, 32(8), 5–9.
- Stevens, S. E. (2009). Is there an ethical duty? Cultural competency. *Oregon State Bar Bulletin*, 69, 9-10.
- Stronge, J. (2006). Teacher evaluation and school improvement: Improving the educational landscape. In J. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice* (pp. 1-23). Thousand Oaks, CA: SAGE Publications.
- Sullivan, C. (2001). *Rewarding excellence: Teacher evaluation and compensation*. Alexandria, VA: National School Boards Association.
- Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge, England: Cambridge University Press.
- Tishman, S., Perkins, D. N., & Jay, E. (1995). *The thinking classroom: learning and teaching in a culture of thinking*. United States of America: Allyn & Bacon.
- Tucker, C. M., Porter, T., Reinke, W. M., Herman, K. C., Ivery, P. D., Mack, C. E., & Jackson, E. S. (2005). Promoting teacher efficacy for working with culturally diverse students. *Preventing School Failure*, 50(1), 29.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*.
- Tyler, K. M., & Boelter, C. M. (2008). Linking Black middle school students’ perceptions of teachers’ expectations to academic engagement and efficacy. *Negro Educational Review*, 59(1/2), 27-44.
- Unesco. (1997). *Appraisal study on the Cyprus education system*. Paris, IIEP.
- U.S. Department of Education. 2010. *Connect and inspire: Online communities of practice in education*. Retrieved April 18, 2012, from <http://edcosp.org/report>
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, Mass.: Harvard University Press.
- Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: The MIT Press.
- Walqui, A. (2006). Scaffolding instruction for English language learners: a conceptual framework. *International Journal of Bilingual Education and Bilingualism*, 9(2), 159-180.
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? a content analysis of review literature. *Journal of Educational Research*, 84(1), 30.

- Wang, J. (1998). Opportunity to learn: The impacts and policy implications. *Educational Evaluation And Policy Analysis*, 20(3), 137-56.
- Waxman, Hersh C., And Huang, & Shwu-Yong L. (1997). Classroom instruction and learning environment differences between effective and ineffective urban elementary schools for African American students. *Urban Education* 32 (1):7-44.
- Waxman, H. C., Hilberg, R., & Tharp, R. G. (2004). Future Directions for Classroom Observation Research. In H. C. Waxman, R. G. Tharp, R. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 266-277). New York, NY US: Cambridge University Press.
- Webb, N. L. (2005). *Web Alignment Tool*. Wisconsin Center of Educational Research. University of Wisconsin, Madison. <http://www.wcer.wisc.edu/WAT/index.aspx>
- Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, 20(1), 7-25.
- Weick, K. E. (2007). The generative properties of richness. *Academy of Management Journal*, 50(1), 14-19.
- Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & ... New Teacher, P. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Second Edition. New Teacher Project.
- Wenglinsky, H., & Educational Testing Service, P. r. (2000). *How teaching matters: Bringing the classroom back into discussions of teacher quality*.
- Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology*, 89(3), 411-419.
- Wertsch, J. V. (1991). *Voices of the mind: A socio-cultural approach to mediated action*. Cambridge, MA: Harvard University Press.
- Western, S. (2008). *Leadership: A critical text*. Los Angeles: Sage Publications.
- Westgate, D., & Hughes, M. (1997). Identifying “quality” in classroom talk: An enduring research task. *Language and Education*, 11(2), 125-138.